

Towards a Theory of Conscious Art

Robert Pepperell

Contents:

- Introduction
- 1. Unconsciousness and self-consciousness
- 2. The inconceivability problem
- 3. Is infinite regression part of the problem or the solution?
- 4. The self-reflecting mirror
- 5. Zen and Tao
- 6. Nen and reflection
- 7. Video feedback
- 8. Reflexive theories of consciousness
- 9. Infinite regression and conscious experience
- 10. Further analogies
- 11. Possible objections
- 12. First steps towards a self-conscious work of art
- Conclusion

Biography

Robert Pepperell is an artist and writer. He studied at the Slade School of Art and went on work with a number of influential multimedia collaborations including Hex, Coldcut and Hexstatic. As well as producing experimental computer art and computer games he has published several interactive CD-Roms and exhibited numerous interactive installations including at the Glasgow Gallery of Modern Art, the ICA, London, the Barbican Gallery, London and the Millennium Dome, London. His book *The Posthuman Condition* was first published in 1995 and is shortly to be published in a new edition, with the subtitle *Consciousness beyond the brain*. His second book *The Postdigital Membrane* was a collaboration with Dr Michael Punt published in 2000. He has spoken and lectured widely on art, philosophy and new technology and is currently a senior lecturer in Contemporary Art Theory at University of Wales College, Newport, a regular reviewer for the Leonardo journal, and Director of the newly founded Posthuman Laboratory for Arts Research (PoLAR). A new book, *Extended Being*, is in preparation.

Abstract

In this paper I argue that when we try to describe the specifically self-aware part of the mind, as opposed to the host of unconscious psychic activities, we face a potentially fatal difficulty—one I have termed ‘the inconceivability problem’. Because of the entanglement of the subject and the object in observations of subjectivity, and certain conceptual circularities, it seems we might never be able to represent the self-conscious mind with anything other than itself. This could leave consciousness studies in a very awkward position. In an attempt to address this I propose that the concept of infinite regression, which is normally associated with the ‘homuncular fallacy’, be reinterpreted productively, in a way that puts self-reference at the heart of our conception of phenomenal experience. Looking at several examples of self-referential systems and theories of mind, including Zen, it seems one system in particular—video feedback—offers a rich source of analogies that might help us to visualise, if not explain, the operation of ‘world-embedded’ self-consciousness. I explain that this inquiry is an attempt to build a theoretical foundation for the construction of a ‘conscious art’, by which I mean a type of art that is, to some extent, aware of itself and its surroundings.

Introduction

This aim of this paper is to suggest a viable theoretical foundation from which it might be possible to construct a work of conscious art, by which I mean a work of art that is, to some extent, aware of itself and its surroundings¹. Some might regard this project as overly ambitious, or even foolhardy, and I have to admit at this time I have little concrete idea of what a piece of conscious art would look like or do. Nevertheless, in what follows I hope to establish sufficient theoretical grounds to justify further practical consideration of what is at the moment not much more than a strongly held intuition: that a conscious work of art would have a potential depth and richness of semantic significance as great as, or even greater than an 'inert' work of art, but would also radically alter the conventional relationship between (unconscious) art object and (conscious) human viewer. This paper does not directly address a theory of conscious art as such, nor the detailed mechanisms through which it might be constructed, but rather the general philosophical foundations on which it might be based. The working hypothesis explored here is that the existence of consciousness, and self-consciousness in particular, owes something to irreducible self-referential processes. Moreover, such processes might be infinitely regressive, where infinite regression is understood as a productive trope as distinct from the way it is sometimes regarded in philosophy of mind, as a logical fallacy.

Like many of us, I've often found myself struggling with the idea of consciousness and reeling under the effects of questions like "How is it that I can think?" and "How do I know I exist?" The traditional philosophical approach to such questions has produced an incalculable number of types of response. I could start the list of 'isms I have personally sought to understand—dualism, materialism, cognitivism, eliminativism, panpsychism, functionalism, behaviourism, idealism—and continue for some time. Yet it seems odd, not to say paradoxical, that philosophy, which is ostensibly the enquiry into objective truth and one of the oldest intellectual pursuits, has so far produced very little on which we can all agree. In fact, we'd be hard pressed to point to *anything* that *every* philosopher can agree on (particularly in the field of consciousness studies).

There is one phrase, however, that seems to invite frequent suspicion and occasional abhorrence in philosophical circles, particularly when used in the context of philosophy of mind—namely 'infinite regression'. I'd like to look again at the notion of infinite regression in the context of consciousness studies to see if it can offer any insights into our understanding of self-conscious experience, and whether it can contribute to the theoretical foundations of a conscious art.²

1. Unconsciousness and self-consciousness

Even since before Freud's and Helmholtz's researches it has been recognised that much, if not most psychic activity is either unconscious or sub-conscious, which is to say it functions without the direct awareness of the subject. Considerable evidence has recently accumulated to support this idea, particularly

of the widely discussed implications of intra-operative memory (Bonebakker et al 1996), blindsight (Weiskrantz 1986), readiness-potential (Libet et al 1983), and size deception (Goodale et al 1995), not to mention the various kinds of hypnosis, suggestion and auto-suggestion which are standard subjects of psychological research. One could also cite those levels of psychic activity more directly implicated in the sympathetic and parasympathetic autonomic nervous systems where the mind and the body imperceptibly merge, the actions of various chemical messengers like dopamine and serotonin on our emotional states and behaviour, as well as the structure of long term memory, habituation, involuntary reflex, and so on (Thompson 1993). What these and many other processes seem to confirm is the depth of those dark pools of desire, impulse and belief to which we as conscious subjects are largely oblivious but which collectively have a direct bearing on, perhaps even a determining influence on our mental being and actions. To distinguish 'consciousness studies' from more general studies of psychic activity one might be tempted to put these unconscious processes to one side and focus on the relatively confined case of the self-conscious, reflexive part of sentience we associate with immediate awareness, active memory, alertness and verbal communication. In other words, the experience in which we know we exist and in which we are able to evaluate and describe how we feel.³

The puzzle of consciousness studies, as many people have described it, is how to account for this "phenomenal concept of mind" which is "characterised by the way it feels" (Chalmers 1996). It presumably includes our 'knowing' what we feel and our 'knowing' our own presence—the sort of awareness, in fact, that gives rise to questions like those in the introduction to this paper as well as all those qualities of lived experience like 'greenness', 'bitterness' or 'tunefulness' we all believe we share. It is these very vivid experiences that seem so remarkably absent within the greyish neurological strata of our brains when we look for them. The problem is not just to place the riot and colour of human mental experience within our silent and monochrome bio-fabric, but to show how they are experienced as being there—as we know them to be.

2. The inconceivability problem

Part of the difficulty of the problem of self-consciousness is as much to do with conceptual circularity as scientific methods. In attempting to objectively study the most self-aware part of our own experience we find strict objectivity becomes immediately impractical. Objectivity requires a distance from the subject, yet in the case of the study of my own self-awareness the object is identical to the subject, and vice versa. To avoid this paradox one might attempt the extrinsic study of other subjects using something like fMRI scanning, analysis of introspective reports, or even psychoanalysis. But here it becomes clear that the subject of study is also the object whose unique experience can only be indirectly and partially examined. By definition an observer cannot remotely share the unique experience of another subject (even with some fantasy mind 'transplanter') if only because the observer could not hold the subject's

experience both separately from and identically with their own.⁴ In short: *There does not seem to be any way to describe, represent or think about the very faculties with which we describe, represent and think, other than with those faculties themselves.*⁵ This is what one might call the ‘inconceivability problem’: like trying to use a hammer to bang itself into a nail, or a camera to photograph the emulsion on its own film; in either case there is a circular self-reference that renders the operation absurd or inconceivable. Should we conclude that any attempt to produce an objective description, representation or explanation of our subjectivity would also lead to absurdity, as it would suffer the same paradoxical circularity? I suspect this is something many in the field of consciousness studies would want to resist, if only on the grounds that it would endanger the whole enterprise.

3. Is infinite regression part of the problem or the solution?

Perhaps the fact that we hit the buffers of circularity so frequently when conceptualising the nature of experience should tell us something? So many of the exotic creatures of recent theoretical writing seem to be haunted by inconceivability: Turing’s universal machine, Nagel’s bat, Searle’s Chinese room operator, Jackson’s red rose, zombies, and all the various numbskulls and homunculi of philosophical literature in one way or another highlight the problem of the displacement of the phenomenal observer to an ever-receding position—the so-called *homuncular fallacy*—leading to a dissatisfying infinite regress. The question has been posed many times: who (what) is it that *finally* looks at our sensory representations of the world and experiences their meaning?

It is a problem that was characterised (although not necessarily resolved) by Daniel Dennett when he referred to “the illusion of the Central Meaner”, the executive mind or “internal Boss” who directs and co-ordinates diverse psychic operations in order to maintain a singular sense of self (Dennett 1991). Speaking of why the mind seems to be ordered in this way, he says “we persist in the habit of positing a separate process of observation (now of inner observation) intervening between the circumstances about which we can report and the report we issue—overlooking the fact that at some point this regress of interior observers must be stopped by a process that unites contents with their verbal expression without any intermediary content-appreciator.” (Dennett *ibid.* p 320). No doubt it is the commonly held expectation of a unifying agent, or conclusive ‘meta-mind’, which leads to philosophical frustration precisely because it remains so elusive. But despite the many theories of experience so far constructed there is no obvious sign of release from the centripetal forces of the regressive homunculi.

As an alternative, therefore, could we not harness this irresistible force productively within our concept of mind such that, in a suitably circular way, it actually resolved its own problem? In the following sections I’ll tentatively examine this possibility, first by looking at some examples of regressive and paradoxical self-reference in different contexts.

4. The self-reflecting mirror

Can a plane mirror reflect itself? The obvious answer is no, but the situation becomes more complex if we introduce one mirror to another where, in effect, a mirror can reflect a reflection of itself. We know from basic physics that two perfect plane mirrors held in parallel (at 0°) reflect each other into 'infinity' so the photons travelling at a right angle to the planes will bounce back and forth 'forever'. In practice, mirroring infinity isn't quite as mind-blowing as it sounds since real mirrors are less than perfect and light is eventually dispersed through deflection or absorption. Nevertheless, we know from our own experiments with mirrors that although the recursive image quickly becomes green and murky, and the head of the viewer inserted between the two planes trying to see eternity itself becomes an inevitable obstruction to the view, there remains a fascination with the image tunnelling into the mist of infinity. And even despite the slightly disappointing behaviour of imperfect reality in comparison to the dazzling promise of ideal conditions, we have to recognise that the two mirror planes offer us an everyday physical example of that which, in conceptual terms, is often associated with fallacy: with two mirrors we create a tangible instance of 'infinite regression' without sending logical shock waves through the universe.⁶

5. Zen and Tao

Although something of the paradoxical character of our conception of self-consciousness, and its inconceivability, has been noted by recent thinkers (Bermúdez 1998, Nagel 1998), there exist much older practical theories of mind which also address the absurdly self-referential and the logically incomprehensible, notably Zen and Tao. According to many of those who have tried to make Japanese Zen thought, and its intimate cousin the Chinese philosophy of the Tao, accessible to those outside the traditions there are aspects of conscious experience that remain forever unutterable. Speaking of what is perhaps one of the most famous expressions in east Asian literature, Alan Watts points to some of the many possible interpretations of the opening line of the *Tao Te Ching*: "The Tao which can be spoken is not the eternal Tao." This translation, Watts continues, "conceals the fact that the ideogram rendered as 'to be spoken of' is also Tao, because the word is also used with the meaning of 'to speak' or 'to say' . . . Literally, the passage says, 'Tao can be Tao not eternal [or regular] Tao.'" (Watts 1976). Just as the self-reflecting mirror is a recurring image in Zen teaching, so the seemingly contradictory self-referential statement is foundational to the Tao, not just as a pedagogical tool but as an act of mental resistance to binary logic and final cause.

6. Nen and reflection

In his excellent book *Zen Training*, Katsuki Sekida (1985) outlines a theory of immediate consciousness using the behaviour of mental actions called 'nen', approximately translated from Japanese as 'thought impulses'. Despite its

relative simplicity I couldn't fully represent his theory here and would urge the reader to consult his book directly where the ideas are set out with admirable clarity. For the purposes of this paper I want simply to sketch the basic principle of nen-action, introduced by a passage from the book itself:

Man thinks unconsciously. Man thinks and acts without noticing. When he thinks. "It is fine today," he is aware of the weather but not of his own thought. It is the reflecting action of consciousness that comes immediately after the thought that makes him aware of his own thinking By this reflecting action of consciousness, man comes to know what is going on in his mind, and that he has a mind; and he recognises his own being. (Sekida 1985)

According to Sekida, thought impulses rise up all the time in our subconscious mind, swarming about behind the scenes, jostling for their moment of attention on the 'stage'. Of these 'first nen', as Sekida calls them, most go unnoticed and sink back into the obscurity of the subconscious, perhaps to return later in some harmful form. But those that are noticed, just momentarily, by the *reflecting* action of consciousness (the 'second nen') form part of a reflexive sequence that supports our sense of self-awareness. The second nen follows the first so quickly they seem to occur simultaneously—they seem to be one thought. The obvious problem of how we know anything of this second nen is resolved by the action of a third nen which "illuminates and reflects upon the immediately preceding nen" but "also does not know anything about itself. What will become aware of it is another reflecting action of consciousness that immediately follows in turn", and so on. Meanwhile, new first nen are constantly appearing and demanding the attention of the second nen.

For the sake of simplicity this sequence is initially presented as a linear progression, but Sekida goes on to elaborate the schema with a more subtle, matrix-like organisation while the basic principle remains. What follows from this is that, as Sekida states, "Man thinks unconsciously"; there is no localisation of conscious thought, no conscious object as such, other than an ongoing loop of self-reflections.⁷ Nevertheless, because of the rapid sequencing of the internal reflections, one has the impression of a sensible self much in the way that one has the impression of moving objects in the cinematic apparatus.⁸ There is an obvious analogy with the self-reflecting mirrors in which the regressive image can exist in neither mirror alone, just as no nen is conscious in its own right. This theory would suggest that the notion of the 'self' does not exist outside a process of continuous self-reflection, nor in any part of that process.

7. Video feedback

The model of self-consciousness based on recurring internal self-reference has strong parallels not only with the self-reflecting mirrors above, but also with another phenomenon familiar to ex-VJs like myself.⁹ It has been known for some time that a video camera pointing at a TV screen can, under certain conditions,

produce startling visual effects of great complexity and beauty, a technique known as 'video feedback' (Crutchfield 1984). To achieve the full effect the most important condition to satisfy is that the camera views its own output signal being displayed on the screen. Within this there are a number of variable parameters that alter the behaviour of the image:

For extreme parameter settings, such as small rotation, low contrast, large demagnification, and so on, equilibrium images are typically observed. For example, when the zoom is much less than unity then one observes an infinite regression of successively smaller images of the monitor within the monitor within . . . The image is similar to that seen when two mirrors face each other. With a bit of rotation the infinitely regressing image takes on an overall "logarithmic spiral" shape that winds into the origin. (Crutchfield 1984)

Although there are many galleries of video feedback images on the Web, what is not so clear from video captures is the magnificent motion of the images, on one hand stable and fluid and on the other jittery and chaotic.¹⁰ Let's briefly list some of the attributes of video feedback, many of which are discussed in Crutchfield's original paper, in the context of the study of qualitative dynamics:

- Video feedback is an exemplary instance of emergent complexity being a product of both local and global conditions.
- It requires the co-operation of a number of sub-systems, none of which can produce the effect alone.
- The image is self-referential as it feeds off its own signal (output and input merge).
- It displays both irregular and periodic behaviour simultaneously.
- It spontaneously generates regions of similarity and difference.
- Overall behaviour is non-linear and subject to a number of variable internal system parameters.
- Any attempt to intervene in the system, or observe it 'from within', will disturb it.

In many ways these attributes of the video feedback image are also necessary general conditions, or analogous properties, of human consciousness:

- Certain cognitive functions, such as memories, seem to be distributed both locally and globally in the brain (Draaisma 2000).
- Human consciousness requires the co-presence of a number of functional systems, and not just a brain, or part of a brain.
- The sensory system and the environment give rise to feedback loops upon which the development of the conscious system is dependent (Edelman 1992).
- Thoughts can occur unpredictably but the mind normally also displays levels of overall stability.

- We are able to perceive similarity and difference at the same time.
- The brain is a non-linear system, highly sensitive to neurochemical and other local and global parameter changes.
- It doesn't seem possible to objectively observe or intervene in the self-conscious mind without necessitating re-interpretation or creating disturbance (see section on 'the inconceivability problem' above).

These latter statements are highly general and are offered here by way of analogy only. But taken together with the theory of non-action proposed by Sekida, perhaps the illustrative case of video feedback offers some useful clues about how human self-awareness might be visualised, if not explained.

If Sekida is correct, and that what appears to be the 'stream of consciousness' is in fact the internal reflection of unconscious thought impulses that are re-reflected ad infinitum, then one physical system which can display this self-reference with a certain visual grace is video feedback. Along with the parallel mirrors, video feedback is another example of how a *relatively* simple physical system with regressive properties does not spin off into conceptual oblivion but, on the contrary, can produce behaviour of great variety, intricacy, and beauty. This variety, intricacy, and beauty is in part generated by the complex, non-linear properties of the feedback system—the chromatic aberrations and distortions of the screen, the varying voltage gains of the red, green and blue channels, the Moiré patterns generated by the discretization of the monitor's phosphors and the camera's charge-coupled device, and so on. Minor local instances of distortion and interference multiply to produce global effects that cannot be attributed to any particular component or aberration. Hence, video feedback demonstrates how self-referential regression in a complex, non-linear system can generate novelty and pattern. If such behavioural richness can be generated in a system as relatively simple as video feedback, then there is even greater potential for it to be generated in the vastly more complex nervous system, which we know is profoundly non-linear.¹¹

8. Reflexive theories of consciousness

Reflexive, self-reflecting or feedback-based models of self-consciousness, of course, are nothing especially novel in recent Occidental theories of mind. For Lacanian psychoanalysis the 'mirror stage' in childhood development is formative of the self-referential *I*; a theory based on observation of the pleasurable reactions of small children and certain apes when presented with their own mirror image (Lacan 1977). In *Gödel, Escher, Bach* Douglas Hofstadter (1980) discusses the application of what he terms 'Strange Loops' or 'Tangled Hierarchies' in modelling human thought and consciousness. Speaking from a position deeply rooted in AI research, Hofstadter draws analogies between the modes of recursion, self-reference and emergent complexity found in the works of Gödel, Escher and Bach and the symbolic interaction of 'subsystems' or 'subbrains' in the production of mind. He concludes:

My belief is that the explanations of “emergent” phenomena in our brains—for instance, ideas, hopes, images, analogies, and finally consciousness and free will—are based on a kind of Strange Loop, an interaction between levels in which the top level reaches back down towards the bottom level and influences it, while at the same time being itself determined by the bottom level. In other words, a self-reinforcing “resonance” between different levels—quite like the Henkin sentence which, by merely asserting its own provability, actually becomes provable. The self comes into being the moment it has the power to reflect itself. (Hofstadter 1980 p. 709)

More recently Gerald Edelman (1992) has been developing theories of “neural Darwinism” or “Neuronal Group Selection” based on principles of competitive selection between groups of neurones, or “maps”, that learn by becoming predisposed to communicate with other maps depending on feedback between the organism and its environment. Central to Edelman’s theory is the parallel bi-directional interaction between maps that allows “re-entrant” signalling and gives rise to a massive system of internal looping which generates huge complexity. Re-entrancy not only enables constant modification and adaptation to new conditions but leads, ultimately to self-representation and “higher order consciousness”, or “consciousness of consciousness”.

More recently still, Francis Crick and Christof Koch have proposed the ‘Unconscious Homunculus’ as a possible model for internal representations of conscious states (Crick and Koch 2000). As noted above, the homunculus has often been equated with the fallacy of infinite regression because it implies that responsibility for phenomenal experience is deferred to agents beyond the observed system and thus leads to the sort of frustrating regression that theorists try to avoid. Taking their lead from Karl Lashley’s 1956 declaration that “No activity of mind is ever conscious” and Fred Atneave’s 1961 commentary entitled “In Defense of Homunculi”, Crick and Koch speculate that the homunculi needn’t themselves be ‘regressive’. By this I presume they mean that the presence of a homunculus doesn’t have to imply the existence of an infinite number of further homunculi observing each other in some indefinite Russian Doll-like arrangement. They conclude in a way that bears similarity to the theory of mind proposed by Sekida above, albeit using a completely different technical language:

As has often been assumed, we are not directly aware of the outer world of sensory events. Instead, we are conscious of the results of some of the computations performed by the nervous system on the various neural representations of this sensory world . . . Nor are we aware of our inner world of thoughts, intentions and planning (that is of our unconscious homunculus) but . . . only of the sensory representations associated with these mental activities. (Crick and Koch 2000)

While Crick and Koch are keen (mistakenly, in my view) to attribute “the subjective world of qualia” to specific brain states, they also recognise the same essentially unconscious self-referencing of one part of the mind by another that Sekida describes in the theory of nen-action.

So, a model of consciousness predicated on self-reflection, self-representation, internal looping or re-entrant signalling would seem to be respectable—perhaps even orthodox—and what’s more, generally consistent with both current science and older theories of mind.¹²

9. Infinite regression and conscious experience

What then can we say of ‘infinite regression’ and the spiralling homunculi of phenomenal consciousness? Is the brain, with its many billions of potential interconnections and internal feedback loops, so inherently complex as to be in effect infinitely self-referential—and hence regressive? Could a regressive, non-linear biochemical process produce experiential novelty and self-awareness, as was suggested by analogy with video feedback above? Before extending the analogy between a self-referential system and self-consciousness, perhaps we need to look more closely at what we mean by the phrase ‘infinite regression’ given its often troubled history in the field of consciousness studies.

We should recognise that cases of infinite regression can occur in both the conceptual and physical domains, where one is a mental construct and the other a material process (which is not to say that mental constructs are not also material processes). For example, the homuncular fallacy is a conceptual instance of infinite regression and video feedback is a physical instance. While conceptual and physical cases are not necessarily identical in character, they are nevertheless inherently similar in being self-referential or recursive. The extent to which one can justifiably transfer the concept of infinite regression between these two domains will be addressed in the section dealing with objections below, but we are now perhaps in a position to recognise the essential ambiguity of its meaning:

The word ‘infinite’ is sometimes used in an absolute mathematical sense and sometimes in a relative sense to refer merely to what is impractical, inconceivable or lacking a clear beginning or end. (For example, it would be impractical to count the number of references to ‘sex’ on the Internet, although strictly speaking the total number would not be infinite.) Either way, this ambiguity often inheres when the phrase ‘infinite regression’ is used in different contexts, particularly when it is transferred between the conceptual and physical domains: is there really an *infinite* number of mirror images in the self-reflecting mirror, or is the number merely indeterminate? To add to the confusion, the concept of ‘regress’ can either be regarded as an indefinite series which gradually recedes from an origin (like the explanation needed to explain the explanation) or as a self-referential loop—this interpretation being closer to its Latin derivation of ‘going back’ or ‘returning’. The former, serial interpretation is usually associated with conceptual viciousness, philosophical redundancy or fallacy while the latter,

as in the physical case of video feedback considered here, may be richly productive when trying to visualise the mind, if only by analogy.¹³

10. Further analogies

I'd like to briefly suggest some further possible implications of the analogy drawn here between video feedback and consciousness. It was Sigmund Freud who, as far back as 1900 in *The Interpretation of Dreams*, suggested consciousness acted as a "sense organ for the perception of psychical qualities" (Freud 1976)—in other words as a 'sixth' sense. It does seem that self-consciousness has the strange attribute of allowing us to see what we see, hear what we hear, taste what we taste, smell what we smell and feel what we feel, as well as think about what we think. Bearing this in mind (and referring back to the larger project of constructing a 'conscious art' object) I'd like to propose an extended video feedback system that includes a video mixer that merges four distinct sources. Imagine the monitor in the video feedback set-up not only displays the image from a camera—A—but also a 'mix' of three other sources, or sub-signals—B, C, and D—such that all four sources merge in the monitor display and are observed by the camera. In such a set-up the feedback image generated by the looped signal A also incorporates the information from the sub-signals B, C and D. Whichever is the strongest sub-signal tends to have a greater influence on the overall properties of the feedback image. Now imagine that the three sub-signals B, C and D represent sensory data from the world, internal impulses of the body, and states of the unconscious mind respectively, and that the camera represents this "sense organ for the perception of the senses", or internal conscious observer posited by Freud. From this quite simple set-up one can draw surprisingly rich analogies with the operation of human consciousness.

Sensory self-awareness. As many have observed, consciousness is always consciousness of some content, and the only sources of content (as far as I can establish) are objects in the world (apprehended by the sense organs), sensations from inside the body (e.g. pains, tingles, hunger), and mental data (e.g. ideas, memories, thoughts), or combinations thereof. Activity in the world, body and brain can go on quite happily without us being in any way conscious of it, but something special or 'phenomenal' occurs when we do become aware of it. For the purposes of this analogy, think of the video camera as the agent of self-reflection that not only sees the combined data-sources of world B, body C and unconscious mind D (the 'content' selected for prominence by signal strength, or level of excitation) but sees 'itself' seeing them, insofar as its own signal A is fed-back into what it 'sees'. As the signals flow an overall, unitary state is reached that obviates the necessity for any further unifying agent, or homunculi, as the system is self-generating whilst also being infinitely regressive, or self-referential. The same principle might be applied to any reflexive sense, the smelling of smells for example, or the feeling of feelings. Multiplied over the whole sensory system, one might start to speculate how a system with a capacity for self-

awareness of some kind might emerge through having an integrated array of feedbacking self-sensors.

The corporeality of phenomenal experience. In his lecture *Conceiving the impossible and the mind-body problem* Thomas Nagel (1998) rehearses a well-worn fantasy of philosophers of mind. In trying to understand what it would mean to see the inside of a subject tasting chocolate from the outside, as it were, he reaches for some future technology of representation that would lay open before us the “truth” about such phenomenal experiences. Like many philosophers, Nagel reinforces the assumption (for it is still an assumption) that this ‘chocolate experience’ is generated by the brain alone, for this is where he suggests we look to see chocolate being enjoyed. Elsewhere I have argued vigorously against this assumption on the grounds that the vital contribution of the body, and indeed the chocolate itself, can be easily overlooked. I want to stress that any model of human consciousness should be ‘world-embedded’; that is, it should take account of both the role of the body (with all its nerves, hormones, enzymes, feedback loops, and so on) and the effects of environmental events and stimuli. So when we talk about the ‘richness’ of human experience, the pleasures and pains, we should not forget that most of them would be unimaginable in a disembodied brain.

And this raises another vexed question: whether neural states (such as those in the brain) are identical with, or correlated to, phenomenal states of experience. One might respond by asking, first, are certain brain states *necessary* for certain consciousness experiences, such as enjoying chocolate? Given there is well-established knowledge about the effects of drugs and lesions on brain tissue and the corresponding effects on thought and behaviour, one can be fairly certain that there is a necessary relation between the organic condition of the brain and what we feel. But are specific brain states also *sufficient* for phenomenal states, i.e. can the brain produce rich sensual experiences alone? I believe we can be equally certain that it isn’t, and it can’t. For a start, no one has ever shown that a healthy brain can function in isolation, detached from the body, and we know from sensory deprivation studies that extended insulation from environmental stimuli leads to the breakdown of normal brain functioning. It is also clear from many developmental studies that environmental interaction is crucial to brain maturation and the framing of subsequent experience. On top of this the brain itself is famously devoid of sensation, and we know from personal experience how sensations are distributed across the body.

Hence the importance of including data from the ‘world’ and the ‘body’ as well as the ‘mind’ in the extended video feedback analogy described above, on the basis that ‘greenness’, ‘bitterness’ and ‘tunefulness’ are phenomenal experiences that depend on the co-action of environmental stimuli, corporeal sensations and mental contents.

Perceiving continuity and discontinuity. Elsewhere I have argued that nature is neither inherently unified nor fragmented, but that the human sensory apparatus gives rise to perceptions which make the world seem either unified or fragmented

to differing degrees depending on what is sensed (Pepperell 2003). The process of infinite regression in video feedback demonstrates how a complex (non-linear) self-referential system can spontaneously give rise to patterns of similarity and difference. If consciousness is in any way analogous to video feedback it may help us to understand why, in a world that may be neither inherently continuous nor discrete, we are able to experience both qualities.

The binding problem. The feedback-state of the conscious process might have some bearing on the so-called 'binding problem' in which regional neural functions seem to cohere in a unitary experience for the subject. Some of the functional parts in the video feedback system are necessarily non-local (the camera lens must be a certain distance from the monitor) but are also connected by light or, like the brain, electrical conduits. In the case of video feedback, non-local components can give rise to coherent global behaviour that can't be isolated to any part of the system. However, the feedback effect itself can only be observed locally; that is, on the monitor or in the camera eye-piece, despite the distributed nature of the overall set-up. Whatever the confusion or variation might be between the sources or sub-signals B, C and D in the analogy described above, the overall feedback image will retain a certain stability and unitary coherence as long as all the variables stay within certain parameters. This could be likened to the unitary coherence of first-person experience.

11. Possible objections

I recognise that the claims made in this paper might be seen by some as idiosyncratic, not say provocative, whilst at the same time open to a number of valid objections, some of which I shall try to address here, albeit briefly.

With respect to the so-called 'inconceivability problem', one might argue that it makes little sense to say the self-conscious mind cannot be conceived of by anything other than itself, since it is only the self-conscious mind that conceives of anything in the first place. In short, we cannot conceive of what is in itself inconceivable. I think this cuts to the nature of explanation itself, and what we expect explanations to yield for us if they are effective. Without wanting to open the whole complex issue here, I suspect we may have to accept in certain cases (phenomenal conscious perhaps being one of them) that the appropriate use of analogy and metaphor will bring us as close as we're going to get to a total understanding of the phenomenon in question, if only because it can't be represented by anything other than itself. Explanation may be the decomposition of phenomena into comprehensible constituent parts and the establishing of causal chains, but there is no guarantee that all phenomena can be decomposed comprehensibly, nor indeed that there will always be a prior cause that fully accounts for subsequent effects. This is not to advocate a particular brand of mysterianism, but an attempt to recognise the depth of the problem of self-consciousness as it is in all its emergent complexity, rather than as we might prefer it to be: reductively explicable. To answer the objection more directly: if

explanations are a content of our mental experience, then mental experience cannot be a content of explanations.

It might be argued that the case of self-reflecting mirrors described above does not represent an example of 'infinite regression' so much as a rather straightforward physical property of mirrored surfaces acting in relatively closed system. Either there is no conceptual error to be resolved here, or in fact the real conceptual error is to confuse the metaphorical allusion to infinite regress with what is a perfectly explicable physical phenomenon. This objection is interesting in that it points to the distinction between regression in the conceptual and physical domains already touched upon earlier. In the conceptual domain we might consider all sorts of possible regressions, such as Aristotle's Unmoved Mover, that have no physical constraints (other than the constraints placed on the physical substance of the mind), and which can therefore remain unresolved in our imaginations. On the other hand, examples of regression in the physical domain, of which the self-reflecting mirror may be one, must necessarily occur within the constraints of physical laws. These laws (such as the Second Law of Thermodynamics¹⁴) tend to constrain the actions of matter and energy so as to prohibit, for example, the kinds of perfect states of infinitely bouncing light particles described in the self-reflecting mirror example. In which case, such self-reflections are not truly infinitely regressive and irresolvable in the way their conceptual counterparts are often imagined—in effect, nature resolves them. Therefore, one could justifiably argue that the transfer of the term 'infinite regression' from the conceptual to the physical domain would be by way of analogy or metaphor, rather than being a matter of strict equivalence. However, in considering this we are implicitly drawn into the question of the relation between mind and matter, and whether certain constraints apply in the physical domain and not in the mental, or whether indeed there is a valid distinction to be drawn between the two domains at all. Such questions are beyond the immediate scope of this paper. For the purposes of what is presented here, I accept that the regression in the self-reflecting mirrors could be regarded as a metaphorical allusion to conceptual regression. But since I do not classify mental activity as 'non-physical' I would argue there is ultimately no essential difference between regression in the conceptual or physical sense.¹⁵ Each case of regression is conceived by a mind (the recognition of regression is always a mental judgement), while physical laws (as far as we know) constrain each mind that conceives regression—an appropriately circular conclusion!

The inclusion of a Zen description of mind in support of the general thesis is obviously open to various criticisms. In the first place it might seem to be a case of using one theory as evidence for another—a potentially unproductive case of infinite regression. Furthermore, one could raise the objection that the 'nen' model proposed by Sekida is purely subjective, even religious opinion, which is unsupported by empirical data. On top of this, one has to recognise that in all the diverse forms of Zen and Buddhism there exists a huge range of opinions, and interpretations of doctrine, many of which are virtually inexplicable outside the very specific cultural and historical contexts in which they evolved. Why then should one form be any more instructive in terms of the study of consciousness

than another? All these points are valid, and I would certainly not wish to suggest that the brief discussion of 'nen' here provides substantial evidence for any particular theory of mind. I would, however, reject the notion that subjective experience, even opinion, cannot make a contribution to our overall understanding of what is, in any case, an entirely subjective experience. In addition, the often-made assumption that Zen, and indeed Buddhism in general, is a religious doctrine should not pass unchallenged. Many commentators on Zen in particular, state explicitly that the practice of Zen requires no leap of faith in the theological sense employed by Christians or Muslims (or at least no faith greater than that we exhibit when we flick a light-switch, or ignite a Bunsen Burner). Rather, it is often described as a rigorous code of conduct developed with the aim of beneficially transforming one's relationship to the world, as well as offering a highly sophisticated theory of mind. (For an accessible and authoritative account of Zen philosophy see Humphreys 1992). Hence I have included the reference to 'nen' here, not only because I believe it is fascinating and relevant to the overall theme of the paper, but also because I take the view that a greater appreciation of this profound philosophical tradition can only enrich our knowledge in this especially difficult area.

Finally, the analogy between video feedback and consciousness has some severe limitations, some of which might be quite misleading. For instance, video feedback is an electromechanical system whereas human consciousness is (at least in part) neurobiological. In addition, the video feedback system is many orders of magnitude simpler than the human nervous system, and this is even taking into account that we know relatively little about the true complexity of conscious processes, or whether they function in any way that is analogous to video feedback. Also, the same objections can be raised against the application of the term 'infinite regress' in the context of video feedback as were raised in the context of self-reflecting mirrors. Again, all these objections are valid to an extent, but in my view do not necessarily undermine the overall analogy. I would briefly point to another case where analogies have been drawn productively between electronic and biological phenomena—cellular automata. By plotting simple recursive algorithms in a computer-generated virtual space, it is possible to generate relatively complex 'behaviour' in which 'creatures' live, move, die, breed and evolve in an astonishingly 'life-like' way (see Levy 1992). Cellular automata are used productively by biologists and researchers in artificial life to model organic processes, despite suffering the same deficiencies, in terms of being electromechanical and relatively simplistic, as the video feedback analogy suffers with respect to consciousness. I would deal with the objection that video feedback is not a true example of infinite regression in the same way as I dealt with it in the case of the mirror example.

12. First steps towards a self-conscious work of art

Having considered the relationship between phenomenal consciousness and infinite regression in some detail, and having looked at some possible areas of dispute, I would like to end by sketching out, in very general terms, how these

ideas might inform a practical investigation into the production of a self-conscious work of art.

It might be interesting for the reader to know that the ideas presented here originated less from the relevant philosophical literature than from a combination of introspection, personal experience, and artistic enquiry. In particular, the practice of meditation, and examination of its related philosophies, has helped to clarify a number of issues to do with the behaviour of mind and its relation to the body and the world. In addition, whilst using LSD some years ago I experienced vivid recursive patterns of luminous colour, very similar to those seen in video feedback, which triggered an intuition about the self-referential operation of the visual system and, by extension, the mind. It is these experiences, together with the various pieces of interactive art I have produced and exhibited over the years, that have circuitously led me consider how it might be possible to construct an object of art that displays some self-awareness.

Using the principles discussed above, a system is now being designed which combines three sources of data from 1. the external world (with sensors for light, sound, and pressure) 2. the internal state of the system (such as levels of energy, and rates of information flow) and 3. repositories of images, sounds and texts to be activated by rules of association (what one might describe, crudely, as 'memories'). Much as in the extended video feedback analogy described above, these three data sources will be synthesised into an overall system-state, which is then 'observed' by separate sub-system of sensors. This observed state is then fed-back into the overall system-state and re-observed, indefinitely. In this way the system will generate a condition of infinite regress (the phrase is used here with caution) not dissimilar to that found in video feedback, which it is hoped will achieve some overall coherence. At the same time, because conditions will constantly vary in the exhibition space (in terms of audience actions, internal system data states, and associative links with stored data), the global behaviour of the system will be non-linear and unpredictable. This, in a nutshell, is how the system will be designed to work.

However, I should stress that I am not claiming any such system, even if it performed well, would actually be conscious in the same way that we are. Nor am I even claiming it would be quasi-conscious, or yet further, that it would be an accurate model of how conscious processes occur in humans. To claim any of these would not only pre-empt the results of the investigation before it even left the virtual drawing board, but would suggest a far grander purpose than the thesis I have presented here could justify. At best the system might have a rudimentary functional self-awareness.

But even given the obvious limitations, I do expect many more artists to become interested in the creative possibilities of self-aware systems. This is on the basis that such systems will have unique and compelling qualities, including a capacity for producing semantic richness in response to audience behaviour, at the same time as generating a *frisson* of expectation amongst audiences as they apprehend an object that displays, albeit in the mildest of forms, some of the same responsive behaviour they recognise in themselves.

Finally, it is clear that what is being considered here touches on what may well be the most difficult question humans have ever tried to resolve about their own condition. I hope these speculations, and others implied but not included, might help to stimulate avenues of enquiry that might otherwise remain unexplored while provoking new connections between seemingly unrelated ideas. In particular I hope the artistic sensibilities I've brought to the consideration of these deeply complex questions might encourage greater dialogue between disciplines such as art and philosophy that share intellectual concerns, but sadly lack a common intellectual heritage.

Conclusion

I have argued that when we try to describe the specifically self-aware part of the mind, as opposed to the host of unconscious psychic activities, we face a potentially fatal difficulty, one that I have termed 'the inconceivability problem'. Because of the entanglement of the subject and the object in observations of subjectivity, and certain conceptual circularities, it seems we might never be able to represent the self-conscious mind with anything other than itself. This could leave consciousness studies in a very awkward position. In an attempt to overcome this, I have proposed that the kind of infinite regression often associated with the homuncular fallacy be reinterpreted more productively, in a way that puts self-reference at the heart of our conception of phenomenal experience. Looking at several examples of self-referential systems and theories of mind, including the Zen concept of 'nen', it seems one in particular—video feedback—offers a rich source of analogies which might help us to visualise, if not understand, the operation of 'world-embedded' self-consciousness. Infinite regression then, understood in relation to phenomenal experience, may be nothing other than a process of perpetual self-reference, however this might occur within the physical substrate of the human system, the non-linear nature of which can give rise to intricate and novel behaviour. By exploiting the mechanical and analogical properties of video feedback systems, including their inherent complexity and creativity, one can envisage a functional model of a self-referential system that might inform a wider theory of a conscious, or self-aware art.

Bibliography

- Anderson, B. and J., 'Motion Perception in Motion Pictures', in T. de Lauretis, and S. Heath, (ed.) *The Cinematic Apparatus*, (New York, St. Martin's, 1980)
- Bermúdez, J. *The Paradox of Self-Consciousness*, Cambridge, MA, MIT Press, 1998.
- Bohm, D., *Wholeness and the Implicate Order*, London, Routledge, 1983.
- Bonebakker, A. E. et al., 'Information processing during general anaesthesia: Evidence for unconscious memory', *Memory & Cognition*, 24, (1996), pp. 766-776.
- Bonshek, A., *Mirror of Consciousness: Art, Creativity and Veda*, Delhi, Montial Banarsidass Publishers, 2001.
- Chalmers, D., *The Conscious Mind: In Search of a Fundamental Theory*, New York, Oxford University Press, 1996.
- Crick, F. and Koch, C., 'The Unconscious Homunculus', in Metzinger, T. (ed.) *The Neural Correlates of Consciousness*, (Cambridge, MA: MIT Press 2000).
- Crutchfield, J., 'Space-Time Dynamics in Video Feedback', *Physica*, 10D, (1984), pp. 229-245.
- Danto, A., 'Depiction and Description', in *The Body/Body Problem: Selected Essays*, (Berkeley, University of California Press, 1999).
- Dennett, D., *Consciousness Explained*, London, Penguin, 1991.
- Draaisma, D., *Metaphors of Memory: A history of ideas about the mind*, Cambridge, Cambridge University Press, 2000.
- Edelman, G., *Bright air, brilliant fire: On the matter of the mind*, New York, Basic Books, 1992.
- Efron, E., 'The minimum duration of a perception', *Neuropsychologia*, 8, (1970), pp. 57-63.
- Feynman, R., *QED: The Strange Theory of Light and Matter*, Princeton, Princeton University Press, 1985.
- Freud, S., *The Interpretation of Dreams*, London, Pelican, 1976.

- Goodale, M. et al., 'Size-contrast illusions deceive the eye but not the hand', *Current Biology*, 5, (1995), pp. 679-685.
- Hofstadter, D., Gödel, Escher, Bach: An Eternal Golden Braid, New York, Vintage, 1980.
- Humphreys, C., *Zen: A way of life*, Chicago, NTC, 1992.
- Lacan, J., 'The mirror stage as formative of the function of the I as revealed in psychoanalytic experience', in *Écrits: A Selection* (London, Tavistock, 1977).
- Levy, S., *Artificial Life*, London, Jonathan Cape, 1992.
- Libet, B. et al. 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act', *Brain*, 106, (1983), pp. 623-642.
- Nagel, T. 'Conceiving the impossible and the mind-body problem', *Philosophy*, 73 (285), (1998), pp. 337-352.
- Pepperell, R., *The Posthuman Condition: Consciousness beyond the brain*, Bristol, Intellect Books, 2003.
- Searle, J., 'Minds, Brains, and Programs', *The Behavioral and Brain Sciences*, vol. 3, (1980).
- Sekida, K., *Zen Training: Methods and Philosophy*, New York, Weatherhill Inc., 1985.
- Thompson, R., *The brain: a neuroscience primer*, New York, W. H. Freeman and Co, 1993.
- Watts, A., *Tao: The Watercourse Way*, London, Jonathan Cape, 1976.
- Weiskrantz, L., *Blindsight: A Case Study and Implications*, Oxford, Oxford University Press, 1986.
- Wiener, N., *The Human Use of Human Beings: Cybernetics and Society*, Boston, MA: Houghton Mifflin, 1950.

For video feedback pictures, texts and clips see:

<http://www.videofeedback.dk/World/>
<http://home.earthlink.net/~spinninglights/>
http://members.tripod.com/professor_tom/galleries/video/index.html

¹ I make the assumption that 'self-awareness' and 'consciousness', while not being necessarily identical, are somehow implicit in each other.

² It's worth pointing out that 'infinite regression' is a term applied in a number of different conceptual and physical contexts, with varying meanings in each. Certain conceptual cases, such as Aristotle's Unmoved Mover who is the final cause of all other causes, or the indefinite chain of axioms needed to support mathematical propositions, are not of immediate concern here. The distinction between conceptual and physical cases of infinite regression will be addressed later in the paper.

³ The suggestion that one sets aside the unconscious in order to focus on the self-conscious doesn't imply, of course, any essential rupture between them.

⁴ This is not to rule out the existence of shared or consensual subjective experiences, of which the cinema is a classic example. But however absorbing they might be they do not completely efface the differences between individual's experiences.

⁵ There are echoes here of Heisenberg's indeterminacy, or uncertainty principle which, according to most interpretations, states the impossibility of measuring both the co-ordinate and the momentum of a particle at the same time (Bohm 1983).

⁶ For a readable and fascinating account of the strange behaviour of light and mirrors see Feynman (1985).

⁷ I am aware as I write this how sentences and ideas seem to emerge from somewhere already formed, and *subsequently* presented to my conscious mind for selection and editing.

⁸ The widespread belief that the impression of motion in cinema results from 'persistence of vision' has often been challenged, not least by Anderson (1980). What is important here is to note that the experience of movement results from viewing a rapid procession of still images, each slightly different. The parallel with the effect of self-consciousness should be clear: each *nen*, or thought impulse, is in itself unconscious (or still) but when rapidly concatenated with others can create an overall impression of awareness (or motion). There is also evidence to suggest consciousness events are discrete, being parsed into "perceptual frames" of about 70 to 100ms average duration (Efron 1970).

⁹ A "VJ" is a "video-jockey", the visual equivalent of a disc jockey, usually responsible for the visual entertainment in night-clubs and at raves.

¹⁰ Some short video clips of video feedback are also available on the Web sites mentioned in the bibliography.

¹¹ 'Non-linear' systems can't be modelled with linear or first-order equations, but are governed by many complex, reciprocal relationships, or feedback loops. They are sometimes referred to as 'turbulent' or

'dynamical'. I'm grateful to Dr Tom Holroyd of the Yanagida Brain Dynamism Group, Japan and the Center for Complex Systems and Brain Sciences, Florida Atlantic University, Florida, for his advice on some of the ideas raised in this paper and for his excellent video feedback site (see bibliography).

¹² In this respect it is also interesting to note a proposition made by the Maharishi Yogi when explaining the Vedic conception of consciousness: "Consciousness is that which is conscious of itself. Being conscious of itself, consciousness is the knower of itself. Being the knower of itself, consciousness is both the knower and the known. Being both the knower and the known, consciousness is also the process of knowing." It is the "self-referential singularity" of these three qualities which "together are the indications of the existence of consciousness." (Bonshek 2001).

¹³ In fact the more one thinks about it, it seems that nearly all examples of infinite regress are strangely linear and circular, or serial and self-referential, at the same time!

¹⁴ The physical law which states that in any closed system the total amount of energy, or order, gradually decreases.

¹⁵ One might even say that all conceptual activity is physical, but not all physical activity is conceptual.