Multiple visual worlds

‘Seeing’ happens to you when you open an eye. You immediately and quite automatically become visually aware of the scene in front of you, including part of yourself. Ernst Mach’s famous figure (Mach 1870) in his Analysis of Sensations depicts this. You see Mach’s study, bookcases, a window offering a view, Mach’s body (apparently reclining on a sofa), his eyebrow and his moustache. When you open both of your eyes you obtain a pair of views (different, as you can check by blinking) that appear as a single one. The view now looks more ‘real’. When you move your head or body, you obtain a series of views that further increases the vividness. Something similar occurs in pictorial perception. A single photograph may yield a more or less vivid image of a scene, but a pair of photographs is better (say a frontal and a profile view of a face, as in a criminal record), and a full photographic coverage is better still (like the coverage of a scene of a crime as documented by a police photographer). A movie gives you a serial order of pictures and yields indeed a very vivid impression. In all these cases you are firmly convinced that you see a physical scene much ‘as it really is’, only more or less clearly.

That you become aware of a three-dimensional scene even from a single perspective is generally thought to stand in need of an explanation. After all, at any moment the eye samples only a two-dimensional simultaneous presence of radiance. According to the remarkable bishop Berkeley, the observer sees objects not directly, but only through their marks (Berkeley 1709). The so-called ‘cues’ are arbitrary signs (according to Berkeley) that one has learned to associate with certain features of the world. Thus you see anger or shame in a face when the radiant power spectrum shifts towards longer wavelengths. In the modern view it is indeed obvious that cues originate (phylogenetically as well as ontogenetically) from repeated uncontradicted experience. “Evolutionary internalised regularities” originate from an endless repetition of chance events. Berkeley hit the nail on the head. Of course, we have gained an independent understanding of many cues from scientific analysis of the structure of the physical world [the “ecological optics” of Gibson (1950)]—just think of shape from shading, texture gradient, disparity, perspective, contour, and so forth. There is little arbitrary here in that sense.

A problem common to all “shape from shading, contour, texture, motion, ...”, in short “shape-from-x” theories (and thus to all cues) is that they necessarily depend on a—sometimes large—number of prior assumptions. For instance, typical shape-from-shading theories (Horn and Brooks 1989) assume Lambertian surfaces of constant albedo irradiated with uniform, collimated beams, and the absence of interreflections (or multiple scattering) and vignetting (occultation of the source by the shaded object itself). The assumptions cannot be checked on the basis of the available optical structure: they have to be accepted at face value. They might well be wrong. One thus prefers assumptions that are ‘generically true’, that is to say fail only in cases that one would be prepared to consider as ‘freaks’. A biological system will survive quite well (on the average that is—bad luck for the individual) if it fails only in freak cases whose probability of occurrence is nil.

An even larger problem with the shape-from-x theories is that they hardly ever yield a unique solution. They typically yield a—often very broad—family of equally valid solutions. The solutions are related to each other—they all derive from the same cues: I will say that they are related via an ‘ambiguity transformation’. When you have some particular solution and apply all ambiguity transformations to it you obtain the complete
set of solutions. Equivalent scenes (by definition!) yield identical photographs and thus are not different to the observer, at the stimulus level. They differ only to the photographer who remembers the moment of exposure, not to the viewers of the picture.

Why should ambiguity pose a problem? Well, suppose I start with a photograph of a scene. Let's agree to call it the ‘actual scene’. From the photograph I try to reconstruct the scene using all available cues. What I obtain will be an infinite family of scenes, all related by some ambiguity transformation. A percept can be regarded as a solution to some shape-from-x problem. Which scene is the ‘veridical perception’?

You might say: “The actual scene, of course!” But then, the actual scene is by no means singled out among the solutions. If the actual scene is to be called ‘veridical’, then you can't deny this epithet to the infinitely many other contenders. Since you typically cannot characterise the ambiguity, you are quite generally in no position to judge whether the perception of a scene from a photograph (say) is veridical or not. Suppose the photographer suffered a lethal accident, then there is nobody around with first-hand knowledge of the actual scene. Again, the picture might not be a photograph at all, but a work of art from the photorealist school, or even a piece of paper covered with darkish fungus growth in a peculiar simultaneous order. There need not even have existed an ‘actual scene’ at any time! The picture is still a valid stimulus. ‘Veridicality’ in the naive sense is really a nonentity. This poses a serious problem to any psychophysical endeavor in this field. I don't understand why so few people seem to worry about it. I do.

‘Equivalent scenes’ can be very wild indeed. Consider a photograph on the CRT screen. It is an array of a thousand by a thousand pixels (say), each pixel being simply a tiny luminous spot (or a triple of these in the case of colour—it makes no difference). Suppose I move the luminous points out on rays from the eye, adjusting their luminosity according to the inverse square law of photometry. When I move the dots out into outer space they become stars. When I adjust their distances randomly, I can have it such that no pair of them is closer than a few light years from each other. Thus I have created a scene composed out of a dispersed cloud of stars. Yet the observer cannot notice the difference since the retinal illuminance doesn't change. The peculiar stellar constellation is equivalent to the dots on the CRT screen. This is similar to Ames's demonstration of a ‘chair’ made out of disconnected parts (see Itelson 1952, page 26). The other Ames demonstrations like the monocular (ibidem, pages 40, 44, and 46) and binocular rooms (ibidem, page 50) show that box-like interiors can be quite differently structured, yet appear the same. In the binocular Ames rooms, straight lines become curved. Another interesting example is due to Leonardo da Vinci (1651/1989). He remarks that a view of a piece of sculpture can be ‘undone’ by painting the statue in such a way that the shading is exactly ‘neutralised’. The statue will then look like a uniform paper cut-out, a silhouette. Nature has long since invented such ‘counter-shading’ as a camouflage device. Amplifying on this, it should be possible to paint a human body in such a way that an (actual) ventral view will appear like a dorsal view. A cinch to the modern body-paint artist. Thus a human in an anterior and in a posterior view can represent equivalent scenes if you only adjust the albedos suitably.

It is easy enough to carry on and pile one weird equivalent scene upon another. I often see striking examples of this in modern art. I recently saw a photograph of a body-painted nude leaning against a brick wall. The body neatly merged into the wall owing to the body-painted brickwork. When a nude can be mistaken for a brick wall then where's the limit? As an observer you are free to adjust the geometry, material properties, camera position and type, light sources, and so forth. There appear to be very few limits indeed on what is possible. The visual observer is indeed quite free to do all that, because it need only be done in the mind. What you do (your assumptions) is part of your perception of a scene. It is ‘the beholder’s share’ of the percept (Gombrich 1959).
Thus pictures, but also glances of real scenes, allow *multiple visual world interpretations* (see figure 1). The same is true of multiple views and serially ordered views, although the ambiguity groups will tend to shrink as the available cues become richer. We all know that few Hollywood directors will go to the trouble of building real castles or moving the set to the moon. The ambiguities don't go away in the movies. Nor do they in real life.

[Figure 1. The Necker cube figure allows infinitely many interpretations, though only a few (mainly two) appear at all reasonable. Percepts toggle randomly between the two main contenders.]

You may well wonder what perceptions are like? Is a perception like a bundle of possible visual worlds (then the observer apparently abstains from a final decision), or is it like a single visual world (the observer really “sticks the neck out”)? I propose to call the first possibility the *multiple visual worlds hypothesis*, the second possibility is simply common consensus (“folk psychology” for academics), although it is hardly ever explicitly mentioned. I will assume that the issue is empirically decidable. If you don’t grant me that, then you have to accept that both the multiple-visual-worlds and the conventional concepts are equally fictitious. Few people I meet seem prepared to accept the latter consequence.

I think that in a great many cases perceptions are more of the multiple-visual-worlds variety than like the single guess. You don't notice this in the laboratory, because most psychophysical methods *force a unique response*. They simulate the decisions-in-action of daily life. That you don't notice the essential ambiguity of perception in real life is likely due to the fact that you don't need to take decisions on issues on which no actions would be taken anyway. That the multiple-visual-worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain multiple-visual-worlds interpretations explicitly, but merely that they don't necessarily resolve ambiguities when this is not specifically required for some action or decision.

Suppose you drive to some destination and have been told to “stay on the main road”. Inevitably there will be a time when you arrive at a fork, or Y-junction, whose two branches are optically identical except for a reflection about your coasting direction (see figure 2): Do they *look* the same to you? *Obviously not*, for one branch must look like the main road, the other like a side road. That’s what you’re looking for, so that’s what you see. You waver. One branch may appear like the main road right now, the other may look more like it the next moment. When a branch appears like the
main road, it looks as if it makes a less sharp turn and is broader, whereas the other branch makes a decidedly sharper turn and is clearly narrower. As you drive on, you find yourself taking one of the two branches (the only other option is to drive into a tree, say), thus resolving the quandary through a ‘decision-in-action’. Before you took the turn there existed two possible worlds; the action made them collapse into one. When you look at a Necker cube you face a similar binary choice, except that no decision-in-action will ever take care of the collapse. Each decision perversely returns you to the same initial condition. You face yet another binary choice. The Necker cube presents you with an infinite tree of Y-branches, like Borges’s “garden of forking paths” (1970a): “… each time a man is confronted with several alternatives, he chooses one …; [in the garden of forking paths] he chooses—simultaneously—all of them. He creates, in this way diverse futures, … which themselves … proliferate and fork …”. As a result you generate a potentially endless, erratic series of percepts. This is not like real life. In living your life (driving to work, engaging in a sword fight, …) you continually make decisions-in-action that are final (no looking back!) and collapse the multiple visual worlds for better or worse. We survivors seem to do pretty well on the whole. It is as if you traversed an endless tree of Y-branches, like one of Werner Reichardt’s (1961) unfortunate bugs (see figure 3), but unlike the Necker-cube case you are never trapped in vicious loops. Facing the consequences of decisions-in-action has made us (phylogenetically as well as ontogenetically) into what we are. In Schrödinger’s (1964) view, consciousness itself is nothing but this facing the consequences of decisions-in-action.

**Figure 2.** A driver arrives at a Y-junction, determined to “follow the main road”. Both branches are optically identical, but which one is ‘right’?

**Figure 3.** One of Reichardt’s unfortunate bugs (clearly a fake specimen in this sketch. Reichardt ran Chlorofanus) trapped in an infinite maze of Y-junctions. When you look at a Necker cube you’re in a similar conundrum—until you voluntarily jump out of the vicious loop.
Notice that it is entirely possible to have different perceptions in the face of the same optical structure (stimulus). This is due to the fact that perceptions invariably contain a considerable ‘beholder’s share’. The beholder’s share derives from the observer’s visual expertise and prior knowledge. This is a familiar enough phenomenon that perception holds in common with, eg, reading literature. Thus it makes a lot of difference whether you read Cervantes’s work or Pierre Menard’s incomparable fragments of the Quichote. Though Cervantes’s text and Menard’s are verbally identical, the second is almost infinitely richer (Borges 1970b). Indeed, the differences between the two texts are striking in all respects: “The contrast in style is also vivid. The archaic style of Menard—quite foreign, after all—suffers from a certain affectation. Not so that of his forerunner, who handles with ease the current Spanish of his time”. Indeed, examples in many fields abound. Think of the drone of the distant dustcropper in the scene of Cary Grant trying to hitch a ride on a dusty road in the Midwest in Alfred Hitchcock’s “North by Northwest”. When you’ve seen the movie once, the sound will never be the same when you watch a rerun. Nelson Goodman’s (1976) example of Rembrandt’s Lucretia and its perfect copy (are they aesthetically different?) is another interesting instance. [Goodman’s answer is yes. According to him: “A philosopher of art caught without an answer to this question is at least as badly off as a curator of paintings caught taking a Van Meegeren for a Vermeer” (page 99).]

A case that is complementary to that of the Necker cube is that of Leonardo’s wall. Quoth Leonardo (1651/1989, page 222),

“... look at any walls soiled with a variety of stains, ..., you will therein be able to see a resemblance to various landscapes graced with mountains, rivers, rocks, trees, plains, great valleys and hills in many combinations. Or again, you will be able to see various battles and figures darting about, strange-looking faces and costumes, and an endless number of things which you can distill into finely rendered forms. ...”

In the Necker-cube case you have a clear-cut optical structure offering a binary choice; in the case of Leonardo’s wall the optical structure is not so easily summarised and the possible interpretations are boundless. At any time, you’re faced with a junction of infinite branches, much worse than Borges’s enchanted garden. As you watch the wall, you perceive floating shreds of halfbaked perceptions that do not clearly alternate like in the Necker-cube case but are often intermingled and may even seem simultaneously (somewhat) present. You generate an errant trek through an infinite maze. Whereas in real life (negotiating the traffic on the way to work, and so forth) your decisions-in-action may have dire consequences, in the case of the Leonardo wall as in the case of the Necker cube you’re fortunately only freewheeling.

Before I made the decision-in-action at the Y-junction, my percept was undefined (multiple visual worlds); after my decision, it became history, part of my mental makeup. The multiple worlds collapsed into a single one at the moment of the decision-in-action. The percept became operationally defined. In real life the decisions-in-action occur at moments that are forced upon me by the world; there is typically no looking back. It is the flow of time. Quoth Saint Augustine (1961), “All the while the man’s attentive mind, which is present, is relegating the future to the past ...” Decisions-in-action happen to you. The multiple visual worlds continually collapse. No doubt this is the reason why your percepts tend to be well defined most of the time.

You may well ask how it can be that you never experience the really freak interpretations though you know them to fit the optical facts quite well. One answer is that the freak solutions are indeed freaky in the sense of being very unlikely. If you see a ping-pong ball illuminated in a certain way and are (later) shown that you actually looked at a kind of broomstick seen end-on you feel cheated. Indeed, you have a right to feel cheated because this only works if both the viewing direction and the angle of
the illumination are confined to very narrow angular regions. Thus the view was very unlikely (on the assumption that all possible viewing directions and illuminations are a priori equally likely); generically the thing will look like a broomstick. Likewise, in the human figure seen from the front that appears to you as a human figure seen from behind, you again may feel cheated when the hoax is revealed. The body-paint has to be applied expertly, and the viewing and illumination directions have to be dead-on in order for the illusion to work. It is a highly non-generic situation. Similar considerations apply to the Ames chair or the stellar constellation. Thus the set of solutions is somewhat graded by greater or lesser likelihood. Yet you may rest assured that the set of acceptable solutions (by any biologically reasonable standard) is still quite appreciable.

To delineate the ambiguity for given cues is an immense task for theoretical ecological optics. It is indeed very far from being completed. Solutions exist for many geometrical cues and some special cases of shading (Belhumeur et al 1999; Yuille et al 1999), but for preciously little else. It is a priori to be expected that the ambiguities will turn out to be different for the different cues. The ambiguity in visual perception will then be the common core of ambiguities for all relevant cues. It will be decades before useful progress has been made in this area. Fortunately it is possible to venture educated guesses at the outcome. For instance, the sculptor Adolf Hildebrand (1893/1945) (a professional in all matters visual) noticed that bas-relief can hardly be distinguished from sculpture in the round when seen from some distance. He derived a theory of vision in which the group of depth-scaling ambiguities plays a central role. He was indeed one of the first to take the multiple-visual-worlds hypothesis seriously and use it to formulate a theory of perception.

From a methodological perspective the multiple-visual-worlds problem opens a can of worms. Consider some examples:

— Suppose you do an experiment with a number of observers and they all do something different. You will probably not even dream of publishing such a result. It is evidently rubbish! Yet it may well be the case that all observers yield equivalent results modulo an ambiguity transformation. If that could be shown to be the case, then they would have identical responses in so far as these responses were causally related to the stimuli.

— Even worse, suppose the observers clearly divide into two groups. Then you might contemplate publishing the result, yet it need not at all be the case that the observers differed in the sense you expected. They may well have exploited the available cues to exactly the same extent and in the same manner.

— Likewise, if the responses of a single observer are shown to depend on the nature of the psychophysical task, then the conclusion needn’t be that one task yielded ‘more veridical’ results than the other.

— Instead of comparing responses with each other, you might be tempted to compare responses to the actual scene. Although this is indeed the common practice, it makes remarkably little sense because the actual scene is in no way special.

When you think this over, it appears that the literature must be full of mistaken reports and inadequately analysed results. (A blessing I don’t have to be very strict with citations in an editorial! But I’m sure that I contributed to the confusion myself.) That these ideas are indeed very relevant is clear from the fact that cases of prima facie very different but on closer analysis equivalent (under a certain ambiguity transformation) responses are not at all rare. Infinite interpretations (the usual case) can’t be visited exhaustively as in the case of the Necker cube, you (as an observer) may not even be aware of ambiguity. The real thing is the picture before your eyes and you have certain three-dimensional feelings with it. That you entertain multiple possibilities can be shown operationally when you let the picture develop over time in a movie-like
fashion. There will be different degrees of surprise for various developments. You will buy many different movie shreds as ‘natural’, thus revealing the indefiniteness of the initial percept. The problem occurs in real life. It is not just fancy theory or mere philosophising.

Jan Koenderink
Physics of Man, Helmholtz Instituut, Princetonplein 5, NL 3584 CC Utrecht, The Netherlands
e-mail: j.j.koenderink@phys.uu.nl

References
Augustine of Hippo (AD 354–430), 1961 Confessions translated into English by R S Pine-Coffin (Harmondsworth, Middx: Penguin Books)
Belhumeur P N, Kriegman D J, Yuille A L, 1999 “The bas-relief ambiguity” International Journal of Computer Vision 35 33 – 44
Berkley G, 1975 “An essay toward a new theory of vision”, in Philosophical Works, including the Works on Vision (The essay dates from 1709) (1975, London: Dent)
Borges J L, 1970a “The garden of forking paths”, in Labyrinths (Harmondsworth, Middx: Penguin Books)
Borges J L, 1970b “Pierre Menard author of the Quichote”, in Labyrinths (Harmondsworth, Middx: Penguin Books)
Gibson J I, 1950 The Perception of the Visual World (Boston, MA: Houghton Mifflin)
Gombrich E H, 1959 Art and Illusion Part III The Beholder’s Share (London: Phaidon Press)
Goodman N, 1976 Languages of Art (Indianapolis, IN: Hackett)
Hildebrand A, 1893/1945 The Problem of Form in Painting and Sculpture translated into English by M Meyer, R M Ogden (New York: G E Stechert, 1945; originally published in 1893)
Horn B K P, Brooks M J (Eds), 1989 Shape from Shading (Cambridge, MA: MIT Press)
Ittelson W H, 1952 The Ames Demonstrations in Perception (Princeton, NJ: Princeton University Press)
Leonardo da Vinci, 1651/1989, in Leonardo on Painting Ed. M Kemp (New Haven, CT: Yale University Press, 1989; originally published in 1651)
Mach E, 1870/1959 The Analysis of Sensations translated into English by C M Williams (New York: Dover, 1959; originally published in 1870)
Reichardt W, 1961 “Autocorrelation, a principle for the evaluation of sensory information by the central nervous system”, in Sensory Communication Ed. W A Rosenblith (Cambridge, MA: MIT Press)
Schrödinger E, 1964 My View of the World translated into English by C Hastings (Cambridge: Cambridge University Press)
Yuille A L, Snow D, Epstein R, Belhumeur P N, 1999 “Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability” International Journal of Computer Vision 35 203 – 222