What kind of indirect process is visual perception?

Commentary on S. Ullman: ‘Against Direct Perception’ (In Brain and Behavioural Sciences Journal,3, 401-404 1980)

Aaron Sloman
School of Social Sciences,
University of Sussex,
Brighton, BN1 9QN
England
(Now university of Birmingham, UK http://www.cs.bham.ac.uk/~axs)

Introduction: historical note

It is hard to disagree with the main points of Ullman’s paper. Even Kant, by 1781, had pointed out, in opposition to empiricist philosophers, that perception requires a ‘manifold’ of sensory data to be *segmented* (to separate objects), *grouped* (to link parts of the same object), *classified* in accordance with flexible schemata (e.g. dogs, trees and polygons come in many shapes and sizes), and *related* (e.g. spatially, temporally and causally). He argued that perception required a process of synthesis not unlike what occurs in imagination. All of this, he claimed, required a massive contribution of the mind (or brain?), determining the general forms and limits of possible perceptual experiences. Gibson produces no serious rival explanation of these facts of common sense.

We certainly do not perceive things the way physicists tell us they "really" are. The properties and relations we perceive are not those described in quantum theory but abstractions useful for planning, executing and monitoring actions, for recognising individuals and classes of individuals, for forming useful generalisations, and for invoking and monitoring higher-level perceptual processes satisfying these needs. It is hardly disputable that what is perceived is in part determined by inherited abilities (e.g. seeing faces), in part by learning (e.g. seeing your mother’s face, the structure of a flower, or the nuances of a dance). We all know how what we see can also depend on circumstances, such as how tired we are, what we want or expect to see, and the like. Any claim that perception is direct therefore either implies that what physicists tell us about the world is false, or it uses a very peculiar sense of the word ‘direct’, perhaps (as Ullman suggests) merely indicating what Gibson finds interesting and worthy of analysis. (I am not disputing that in other respects Gibson has made very useful contributions to the study of perception.)

In view of all this I find it hard to regard the claim that perception is direct as a serious contribution to psychology, and will therefore restrict myself to minor quibbles over details of Ullman’s arguments and the MIT view of visual perception, after making a small point in partial support of one of Gibson’s claims.

Perceptions and sensations produced in parallel

In section 3.2.1 Ullman mentions Gibson's theory that perceptions and sensations are produced in parallel by different processes. This could be true even if his claim that both are "direct" results of external stimulation is false. Processes of perception can be distinguished from processes of sensation, namely becoming aware of the sorts of things usually referred to by philosophers as "sense-data", e.g. features and relations in the two-dimensional visual field, such as coloured patches and the elliptical appearance of circular objects viewed obliquely, etc. Gibson was in part reacting to philosophers who claimed that perception involves inferences or constructions based on conscious processes of sensation. But normally the latter processes do not occur during perception: for instance if we are not painters or philosophers we may never notice anything elliptical, nor discern acute and obtuse corners, when we see a penny on a table. It requires special training to become aware of the contents of the visual field, as opposed to the contents of the environment.

Thus Gibson was probably correct in saying that perception and sensation (that is, awareness of sense-data) are independent processes, even if he was wrong in denying that either of them requires complex constructive (but unconscious) processes. They are independent only in that each can occur without the other. Of course, granting Gibson this point, not acknowledged explicitly by Ullman, does not undermine Ullman's other criticisms. The independence of the two processes does not rule out their sharing many unconscious "low-level" sub-processes of feature extraction, description, and interpretation. Thus they can be parallel without being 'direct' in any interesting sense.

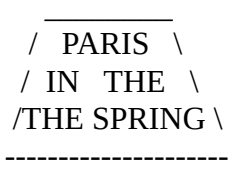
Beware of mathematically tractable special cases

In discussing the recovery of shape from motion (3.2.2), Ullman notes that when the human visual system is presented with a mathematically adequate though impoverished stimulus it will not always perceive the correct structure. He seems to interpret this as due to a failure of the visual system to pick up the available information, and he then launches into a discussion of physiological processes involved in registering properties of the optic array and producing binocular fusion.

But it is possible that the failure of the human visual system to use available information may not be due to a failure to pick up the information. Ullman does not, for instance, consider the possibility that human perception of moving shapes primarily uses mechanisms and strategies appropriate for *non-rigid* motion, such as changing facial expressions, a closing fist, or peel being pulled off a banana. This generally requires more information than rigid motion: and a failure to cope in the situations mentioned may be due to the fact that the mechanisms (or algorithms) require more information, even though mathematically such information is not necessary for the perception of *rigid* motion. Of course, such a system would be able to cope with rigid motion as a special case, when provided with enough information, just as the ability to see curved lines and surfaces may enable straight and flat ones to be perceived as special cases. Notice how few points are required mathematically for "perception" of these special cases: the fact that two points define a straight line may be of no use to a visual system that has to be able to decide whether the line is straight or curved.

So, an adequate analysis of the failure requires a fuller discussion of the difference between failing to pick up information and failing to use it. This note of scepticism concerning Ullman's theory of motion perception does not undermine his discussion of processes by which visual information is picked up. However, I suspect that any account of perceptual processes which can readily be expressed in terms of physiological processes, without the need for higher level 'virtual processors' (see below) would be regarded by a Gibsonian as a theory of 'direct perception'. Stronger anti-Gibson arguments are needed.

One of my favourite anti-direct-perception demonstrations is the well-known example shown in figure 1:



Many people (the exact percentage is irrelevant), when first confronted with this can stare at it for several minutes without seeing anything wrong, despite repeated exhortations to look carefully. The failure to perceive the printed words correctly does not imply that there is any failure in the lower levels of the visual system to pick up the relevant information. (It is interesting that some people discover what is wrong spontaneously if asked to shut their eyes and count the words in the triangle. They often cannot say thereafter which occurrence of "THE" they had previously seen.)

Common observation of human abilities and inabilities suggests that there are many different levels and sub-processes in which things can go wrong, and a study of different sorts of perceptual errors can help to show just how wrong Gibson's theory is. For instance, the 'doubletake' phenomenon (thinking you've seen **X,** then quickly and spontaneously realising it was **Y** after it has moved out of view) lends support to the extended Kantian theory sketched in chapter 9 of Sloman[1978] and Sloman and Owen[1980] that perception involves processing many domains of structure in parallel, with partial results in each domain constraining searches in others.

This organisation partly accounts for flexibility and graceful degradation in difficult circumstances, such as occluding objects, poor lighting, fog, intervening bushes, eye defects, and the like. A theory of direct perception cannot explain such abilities except by vacuous invocation of unspecifiable invariants, and invariant detectors with a magical ability to cope with novel and difficult circumstances.

If the visual system jumps to conclusions on the basis of both partial information and (for the sake of speed) partial analysis at higher levels, this may normally work if the space of possible shapes is sparsely instantiated in the actual world: For example, not all shapes intermediate between a sheep and a horse, or a horse and a giraffe, are found. However, it requires the system to deploy knowledge about which shapes are instantiated, in order to use the redundancy in the optic array. Some sorts of perceptual mistakes suggest that we do indeed deploy such knowledge. But that is inconsistent with any theory that perception is direct.

Dropping out of consciousness

It is curious that Ullman has to rely (in section 5) on Schrodinger's idea that processes perfected in the course of evolution drop out of consciousness. Isn't it a commonplace that many processes perfected through painful individual learning drop out of consciousness -- e.g. reading, playing a musical instrument, sight-reading music, following a spoor in a jungle, driving a car, perceiving botanical or geological structures? To a suitably experienced person, these processes have the same subjective ease and immediacy as the simplest perceptions. The same is true of looking through a peephole at a static scene, where the lack of stereopsis, parallax, and optical flow causes ambiguities about relations between objects which cannot be resolved without prior knowledge. It is quite remarkable how little the absence of these ambiguity resolvers affects our perception of scenes involving familiar objects. Try, for instance, covering and uncovering one eye, repeatedly, with your head quite still. A small peep-hole will help to eliminate information provided by accommodation and head movements.

Ullman seems to grant too much to Gibson. For despite the fact that the optic array in such cases contains an enormous amount of information (if lighting is good, fog and smoke are absent, etc.) it is still inherently ambiguous about occluded parts of objects and relative depths of separate objects. So the fact that we see a specific scene implies that we go beyond available information, contrary to Ullman's claim that ''the role of the processing is not to create information, but to extract it, integrate it, make it explicit and usable''.

Why this refusal to admit that creative inference plays a role in vision? I suspect that it arises out of a desire for theories concerned with mathematically tractable, unambiguous, information-extraction, which in turn is closely bound up with the methodological position Ullman derives from Marr and Poggio. I shall criticise this in the next section.

All this suggests that it is no accident that we find the interpretation of paintings and drawings so easy: infants require no specialised training, because the processing of inherently ambiguous and impoverished information in the light of prior knowledge is a normal part of perception.

These facts seem to be more convincing than the example Ullman offers against Gibson, namely stereopsis (though his point about degrees of directness is a good one). As I've already suggested, Gibson might be happy to describe stereopsis as ''direct'' if based on the sorts of physiological mechanisms indicated by Ullman. Why doesn't Ullman use the more obvious and powerful arguments against direct perception? Is it related to his overall methodological position?

Are there three levels of understanding?

In section 5, following Marr and Poggio, Ullman sketches the methodological assumption that it is important to distinguish three levels of understanding: function, algorithm and mechanism. I think this assumption is confused and fails to acknowledge some important lessons from Computer Science and Artificial Intelligence. Moreover, it threatens to divert attention from difficult and messy problems in psychology to relatively simple mathematical problems.

First of all, the alleged top level cannot be usefully separated from the level of algorithms and the study of representations. For instance, consider the favoured example of pure number theory: for centuries the specification of algorithmic processes (for finding factors, solving equations, and so on) has been central to the theory. That is the source of our concept of an algorithm!

Further, the abstract properties of representations and operations on them have always been central to the theory of numbers, for instance the relationship between representing a number as a sum of powers of 10, a product of powers of primes, a sequence of applications of the successor function, and so on. Even the relationships between these abstract structures and algorithms and the more concrete notation-specific instantiations are very intimate. That is why some philosophers of mathematics have been tempted to analyse mathematics as concerned with nothing but formal manipulations of symbols. We see then that for number theory at least the distinction between the top level and the level of algorithms breaks down completely.

Further, the alleged distinction between algorithm and mechanism fails to take account of the important notion of a ‘‘virtual machine’’. A physical mechanism (e.g. a calculator, or computer, or brain, perhaps) may instantiate a particular virtual machine which can be used as a basis for implementing other virtual machines (using programs which define operating systems, compilers, interpreters, and so on). There can be many layers of different superimposed virtual machines, and the structure need not even be hierarchic (if, for example, a relatively high level program is called as a subroutine from inside the microcode of a computer). Compare Sloman [1978, chs 1,6,10].

Many of the most important issues in AI have been concerned with the study of trade-offs between different virtual machines for a particular function, such as trade-offs between space and time, efficiency and flexibility, efficiency and modularity, completeness and speed, clarity and robustness. It is possible that such computational trade-offs are the key to much of the complexity of human and animal psychology, and ultimately neurophysiology. If so, it may be a serious impediment to scientific progress to advocate an oversimple methodological stance. The calculator example of section 5, for instance, is dangerously misleading, because the rigidity of function of a typical calculator makes it unnecessary for our understanding of it to involve consideration of many layers of implementation or the kinds of trade-offs and mixtures of levels found in human psychology. By contrast, when we study *human* arithmetical expertise (acquired after many years of individual learning), most of the mathematical theory of numbers is an irrelevant digression. Instead we have to consider issues of storing many ‘partial results’, indexing them, linking them to methods of recognising situations where they are applicable, associating them with monitoring processes for detecting slips and mistakes, etc. (Sloman 1978, chapter 7.)

Similar issues arise in the study of human expertise in producing and understanding a natural language: instead of a mathematically elegant formal grammar, a typical speaker seems to use a huge collection of not completely consistent partial rules and heuristics for deploying them. I believe that this is an inevitable consequence of the need for rapid performance, and reliability in circumstances with varying amounts of noise and degradation of sentences produced by other speakers. The same messy kind of complexity would characterise much of visual perception, for much the same reason, even if the lowest levels of the visual system, discussed by Ullman, are an exception, embodying knowledge which can be safely compiled into ‘‘hardware’’ because the physics and geometry of light and many sorts of surfaces are constant in all visual environments. Variable aspects of the environment will need to be dealt with in a different way, mediated by considerable individual learning.

In short, Neisser’s unease with ‘‘processing and still more processing’’, quoted approvingly by Ullman (in section 5), may in fact turn out to be unease with a central feature of human psychology.

Is subjective experience a complete mystery?

In section 5 Ullman claims that experience isa mystery, despite recent attempts to remove the mystery (e.g. Dennett [1979] and chapter 10 of my [1978]), to say nothing of the much older paper by Minsky [1968].

Important steps have been taken by work in AI, showing how in principle internal processes can occur which reflect some of the phenomenological structure of visual subjective experiences -- for example the experience of certain things forming a totality, of one thing being above another, of an edge appearing convex or concave. Of course this work is in its infancy, but it is so far ahead of anything previously available that to say we are still faced with a ‘‘complete mystery’’ is misleading.

For instance, we can now begin to see how other aspects of subjective experiences can be accounted for within the computational/representational approach. The phenomenology of emotional states such as anger, terror or embarrassment requires the use of additional computing concepts, such as priorities, resource allocation, and interrupts. To illustrate: a characteristic of heated emotional states, such as anger or embarrassment, is that attempts to think about something else constantly fail, suggesting that a process of resource allocation is using something like priorities and interrupts. I am currently engaged in a more detailed study of such experiences in collaboration with a research student, Monica Croucher. It is important that in a journal such as this the claim that subjective experience remains a *complete* mystery should not go unchallenged. However, this is not the time for a more detailed discussion. It is worth noting that there will always be a residual area of *moral* disagreement over whether the mystery has been removed, since for example the question whether a robot has subjective experiences is in part a question of how it ought to be treated. Disagreements of that sort, whether concerned with machines, animals or people, cannot be eradicated by science or logic.

References

Dennett, D. C. *Brainstorms,* Harvester Press, 1979.

Kant, Immanuel, *Critique of Pure Reason,* 1781, Translated by N.K.Smith Macmillan, 1929.

Minsky, M L, "Matter mind and models", in *Semantic Information Processing,* MIT Press, 1968.

Sloman, Aaron *The Computer Revolution in Philosophy: Philosophy Science and Models of Mind,* Harvester Press and Humanities Press, 1978.

Sloman A, and D. Owen, "Why visual systems process sketches" in *Proceedings AISB Conference,* ed. S. Hardy, Amsterdam 1980.

Ullman, Shimon, "Against direct perception", in Behavioural and Brain Sciences, (1980)3, pp373-415