

# Visual Conviction

J.A.D.W. Anderson

Intelligent Systems Group, Department of Computer Science,  
University of Reading, RG6 2AX

---

*The state of mind of a fallible, rational, self-conscious, visual observer is considered with regard to the questions: what distinguishes visual knowledge from other kinds of knowledge, and what is the strongest warrant for belief that such an observer can hold? The answer given to the first question is that visual knowledge is any kind of knowledge that the observer could hold in a two-way, spatial mapping with a possible sensory image. The answer given to the second question is that the observer can not have meta-knowledge of any kind, including self-conscious knowledge, but can, at best, hold meta-convictions, where conviction is defined to be consistently justified belief. Consistency checking gives rise to a method for invoking mental strategies by checking their consistency with justified beliefs. Thus consistency checking might bootstrap intelligence.*

---

What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is. (Marr<sup>1</sup> p. 3.)

Marr's answer raises the two major issues of this paper: what can an observer *know* about the world from images, and why is knowing *where* considered to be part of vision? These questions are tackled here by philosophical argument, but they should also yield to sustained technical arguments from Computer Vision and Artificial Intelligence. Thus, this paper establishes my philosophical position on vision, leaving vindication to future work.

## SENSORY KNOWLEDGE

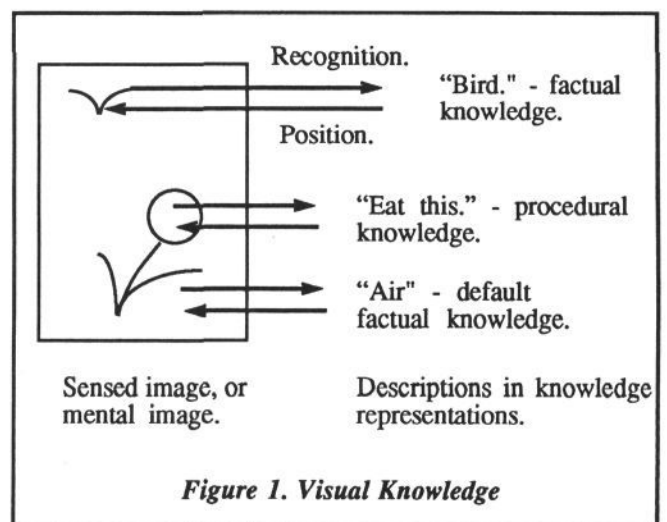
It seems reasonable to talk of auditory, visual, or even general sensory knowledge, but what should these terms be taken to mean? Suppose that someone says, 'There is a bird and a flower in Figure 1'. This would normally be taken to convey a proposition about a figure, a bird, and a flower. The historical fact that someone said this would not make the *proposition* auditory knowledge. The other senses are treated similarly: so that the sensory modality by which knowledge is obtained does not make it knowledge of that modality. I believe, however, that in normal use it is supposed that sensory knowledge mediates at least the bottom-up mapping from a sensory experience to a proposition. Further, it is

normally supposed that a person can re-create a situation so that a sensation is experienced again, which requires a top-down mapping from a proposition to an action. Thus it is normally taken that general sensory knowledge is at least part of a two-way mapping between a sensation and a proposition.

More generally, knowledge representations might be used, rather than propositions, so as to admit procedural knowledge such as skills, in addition to propositional knowledge of facts. Thus sensory knowledge is defined to be at least part of a two-way mapping between a sensory experience and a knowledge representation.

In audition and vision top-down mappings seem to come closer to the stimulus than in other senses. For example, people can produce auditory and visual signals in sophisticated public languages, say, by speaking or drawing, but the senses of taste, smell and touch have less well developed linguistic use, if they have any at all. In the next section, it is argued that the top-down mapping in vision is closer still to the stimulus and involves an observer in the use of a private visual language. Thus visual knowledge properly encompasses a downward mapping, in addition to an upward one, and can, therefore, be defined by a two-way mapping between a sensory experience and a knowledge representation.

## VISUAL KNOWLEDGE



The notion that knowledge is to be described in a computer using some kind of *representation* is familiar, but what is it that distinguishes a *visual* representation from any other sort? Sloman<sup>2</sup> (p. 387) provides a partial answer: '... the difference has to do with whether the representations constructed are closely related to "analogical" representations of a field of view'. So what close relationships, or mappings, might a representation have with a field of view?

A shorthand will be used. Rather than discuss mappings with the field of view, mappings with the sensed image will be discussed. A sensed image is a projection, that is, an analogical representation, of the field of view so nothing is lost by this device and some clarity is gained.

It is evident that vision involves a one-way mapping from the sensed image to what are called knowledge representations. This is necessary for visual recognition. It is more difficult to make the case that vision also involves a mapping from knowledge representations to the image. Straight away, it must be said that this mapping need not exist explicitly. It is conceivable that an observer could manage quite adequately with implicit mappings from visual representations to procedural actions in the world. However, the case will be made that people do in fact possess private mappings from knowledge representations to the image or field of view and that these mappings are spatial in the sense that particular locations within the image map onto knowledge representations, and *vice versa*.

Suppose that an observer is presented with Figure 1 and says, 'There is a bird and a flower in Figure 1'. This would normally be sufficient to credit the observer with a mapping from an assumed sensory image to a knowledge representation. We would be astonished if the observer could not then tell us where the bird and the flower are, perhaps by pointing. To explain this, we might suppose that the observer were being uncooperative or was, in fact, pathologically brain damaged. Such astonishment and a search for an alternative explanation would expose our very strong expectation that people have access to a private, spatial, two-way mapping between an image and a knowledge representation, and can make this explicit in a public language.

The case can be argued technically from Artificial Intelligence. Suppose that an expert system can recognise objects in an image, this necessarily requires a mapping from the image to knowledge representations. Following established methodology it is required that the expert system justify its recognition to the human user. The system's recognition involves image processing of some sort, so it is required to justify its image processing. The only reasonable way to do this is by indicating some region of the image and the corresponding conclusions, perhaps with supporting image statistics. Indicating a region in the image requires a mapping from a knowledge representation to the image and completes the two-way, spatial mapping.

The case can also be argued technically from Computer Vision. A Kalman filter<sup>3</sup> is an optimal method of estimating model parameters which, in theory, may include visual models. It operates in four steps:

1. predicting the occurrence of a target feature in the image from an internal, parametric model;
2. differencing the prediction from the image;
3. feeding the measured difference back into the filter;
4. re-estimating the internal, model parameters using the difference signal, and then looping to step (1).

On every iteration the predicted position of the target feature is optimal, according to the assumptions on which the Kalman filter is built. Thus there is a mapping from a representation, a model in the Kalman filter, to the image, step 1, and a mapping from the image to the representation, steps 3 and 4, which preserves spatial information. Thus optimal estimation of a visual model, by a Kalman filter, involves a two-way, spatial mapping between a knowledge representation and a sensed image. Dynamic Programming<sup>4</sup> and some statistical recognition techniques have similar properties.

Thus there are three independent justifications for there being a two-way, spatial mapping between an image and a knowledge representation: we expect this of people, it is necessary to justify conclusions drawn from an image to another visual observer, such as a person, and it is required by optimal image processing techniques such as Kalman filtering and Dynamic Programming.

Note that the arguments have shown an actual, two-way, spatial mapping, but it is *not* required that a representation is in an actual mapping with an image, only that it could be. Thus a mental image of 'a flower' may be called visual knowledge by virtue of the fact that an observer could map it onto a sensed image, if a sensed image were to contain a suitable projection of a flower.

## KNOWLEDGE OR CONVICTION?

Philosophers have long debated the nature of factual knowledge and have come to the consensus that there are three conditions necessary for a being to know that a proposition holds (see 'knowledge' in<sup>5</sup>). These are that the being holds a justified true belief that the proposition is true. However, Gettier<sup>6</sup> has objected that these criteria are not sufficient to define knowledge, because at least some justifications can be shown to support their conclusions only 'accidentally'. This objection can be met by saying that the justifications must not be open to any defeating objection, that is, they must be indefeasible. Therefore, the currently accepted necessary and sufficient conditions for it to be said that a being knows a proposition are that it holds an indefeasibly justified true belief that the proposition is true.

Now it may happen that all of an observer's beliefs meet these criteria, but if the goal of Computer Vision and Artificial Intelligence is to produce a self-conscious being then, at least in some cases, it must know what it knows, because self-consciousness requires some form of meta-knowledge. Therefore, a self-conscious being must be able to satisfy the criteria of knowledge at least in so far as it has knowledge of itself. However, it is assumed that all observers are fallible, so an observer can not establish beyond doubt that its beliefs are true, nor that its justifications are indefeasible. All that remains of the criteria for knowing is justified belief. Therefore a fallible, self-conscious observer cannot have knowledge about itself, or any meta-knowledge at all, but must accept beliefs with some lesser epistemological status, though not necessarily a status as low as justified belief.

The widely held position that a rational being must hold consistent beliefs is adopted (see 'consistent' in<sup>3</sup>), but keeping in mind that the criteria for knowing were criticised on the adequacy of justification, it is required that every step of a justification (a belief or entailment) is consistent with every belief, including every step of every justification. If the entailment used in a particular program is truth preserving then consistency of justification will follow from consistency of belief, but I wish to admit non-monotonic logics whose entailment may not be trivially truth preserving. However, as a fallible observer is assumed this definition must be weakened to requiring the observer to take it that the criteria hold and not that they actually do hold. This admits the possibility of error at any stage as required by the assumption of fallibility.

Thus 'conviction' is defined to be consistently justified belief and 'visual conviction' is defined to be conviction which is in a two-way, spatial mapping with an image.

For example, Figure 1 indicates beliefs which are in a two-way, spatial mapping with an image. These are that a particular part of Figure 1 could be mapped onto a projection of a bird and that another part could be mapped onto the projection of an edible object (a flower). To be convictions these beliefs must be consistently justified. If an observer were to discover, say, that what it took to be the flower could not be eaten, then it should conclude that there has been some error of observation, either in the original observation or in the discovery, and must amend its beliefs to produce a consistent set. I accept the argument due to Van Fraassen<sup>7dp</sup> that this empirical adequacy is the strongest warrant for perceptual belief, but choose to require the epistemological properties of conviction on pragmatic grounds.

## BOOTSTRAPPING INTELLIGENCE

It is of great practical importance how an observer arrives at a consistent set of beliefs. I can suggest only an evolutionary mechanism. If an observer has the property that it maintains consistently justified beliefs, then it has a mechanism to invoke mental strategies by testing that its beliefs are consistent with it. Thus an

observer might maintain strategies which, say, break the conflict between mutually inconsistent beliefs by adopting a set with high utility. Evolutionary pressure would tend to produce observers whose beliefs, mental strategies, and measures of utility were well adapted to their environment.

This is an empty argument when it comes to demonstrating particular consistency strategies, but a strong one for showing the relevance of bootstrapping epistemologies to Artificial Intelligence.

## CONCLUSION

It has been argued that a fallible, rational, self-conscious observer cannot have knowledge of itself, nor any meta-knowledge at all, but can, at best, hold convictions: consistently justified beliefs. These may be held in a hierarchy of meta-convictions. A conviction is defined to be visual if and only if it may be put into a two-way, spatial mapping with a possible sensory image. Checking consistency gives rise to a method of invoking a mental strategy, by checking the consistency of a belief with it and, therefore, might bootstrap intelligence. Thus the definition of conviction contains the seed of a mechanism for intelligence.

## ACKNOWLEDGEMENTS

I would like to thank Geoff Sullivan and Garfield Dean for many helpful discussions, and Keith Baker for supporting this work.

## REFERENCES

1. Marr, D. *Vision* W.H. Freeman and Company, San Francisco (1982).
2. Sloman, A. 'Image Interpretation: The Way Ahead' *Physical and Biological Processing of Images* eds. O.J. Braddick & A.C. Sleight, Springer Verlag, Berlin (1983) pp 381-400.
3. Kalman, R.E. 'A New Approach to Linear filtering and Prediction Problems', *Trans. A.S.M.E.* 82 (1960) pp 33-45.
4. Sleight, A.C. 'Segmentation and concatenation of edgel lists by dynamic programming and stochastic models' *Proceedings of the Fourth Alvey Vision Conference*, University of Manchester, (Sept 1988) pp 287-296.
5. Laurence Urdang Associates Ltd. *A Dictionary of Philosophy* Pan Books, London (1979).
6. Gettier, E. L. 'Is Justified True Belief knowledge?' *Knowledge and Belief* ed. Griffiths A.P, Oxford University Press, Oxford (1967) pp 144-146.
7. Van Fraassen, B.C. *The Scientific Image*, Clarendon Press, Oxford (1980).

