An Account of Image Perceptual Understanding Based on Epistemic Attention and Reference

Nicolas J. Bullot
University of British Columbia, Department of philosophy, 1866 Main Mall E370,
Vancouver BC, V6T 1Z1, Canada
nbullot@interchange.ubc.ca

Abstract
Technological and scientific images, and other images with epistemic uses, have varied appearances and functions. They seem to be analog or symbolic representations available to researchers for a variety of epistemic purposes such as summarizing data, or presenting, discussing and verifying hypothetic propositions about the world. This article studies the perception and understanding of scientific/epistemic images within a conceptual framework grounded in the notion of reference. It introduces the hypothesis stating that the performance of the perceptual understanding of a particular scientific image depends on the epistemic uses of attention. The hypothesis suggests that understanding a scientific picture requires making an epistemic use of the attentional control of visual routines in order to obtain knowledge on the spatial structure and the referents of a particular image or graphic representation.

1. Phil. & CS: The cognitive problems of image/depiction understanding

Technological and scientific practices rely routinely on the perception of technical tools such as optical systems and measure instruments. By means of these tools, and sometimes without them, theoreticians record data which are usually presented and communicated via a variety of images and graphics such as drawings, photographs, diagrams, mathematical graphs, schemas, IRMf scans, 3D models etc. One can call this variety of non-linguistic but epistemic representations '**scientific images**' or '**pictures for epistemic uses**.' Scientific images are representations available to researchers for a variety of epistemic purposes such as summarizing data, or presenting, discussing and verifying hypothetic propositions about the world. An important question for the epistemology of scientific images is the following:
How are we to describe and explain the semantic and psychological conditions of the epistemic uses of images – as opposed to their socio-political or aesthetic uses? This problem can lead to different types of semantic investigations in cognitive science:

2.1 Two sets of problems

(PB-S1) A **first set of problems** bears on **the structure or conventions of the system which endow the image with its meaning** (e.g., a 'pictorial system' in Lopes' sense), and determine the competence required for its grasping: How does a pictorial system obtain its scientific use(s) and its meaning within the scientific community?

(PB-S2) A **second set of problems** encompasses the questions about **performance of image producing and understanding**. How does an agent understand the meaning of a scientific picture (when he/she is producing it or examining it) in the real time examination or production of the image? How do we use our sensory and motor apparatuses for producing or understanding a scientific or a technical picture?

2.2

(**On PB-S1**) Studies about the first set of questions have been introduced namely by Nelson Goodman [1], Dominic Lopes [2] and Keith Stenning [3]. It has lead to the formulation of conceptual debates about the structure of a pictorial system.

(**On PB-S2**) Although fundamental for the semantic of pictures, these analyses do not answer directly to the second group of questions which bears on the actual cognitive performance of image understanding (instead of the semantic system which is required for understanding pictorial systems). The analysis below intends to show that **resolving the second set of question requires a theory of epistemic attention** within a **referentialist framework** for the semantics of image perception. I will first sketch the referentialist framework and then propose a role for epistemic attention within this framework.

3. A referentialist framework for the semantics of scientific images

This article will suggest a hypothesis which is dependent on a **'referentialist' framework** for the cognitive semantics of image uses.

Referentialist framework: I intend to convey the idea that image understanding is **based on the knowledge of the referents** that parts of images' surface inform about or denote. In the most straightforward case, a referent of a particular image part *e* is any object – or part of an object – that is represented or denoted by *e*.

Scientific images refer to (or represent, depict) **many kinds of referents** such as individual objects, spatial structures, abstract types or variations among magnitudes.

4.1 Hypothesis

General formulation:

***H*, Image Understanding through Epistemic Attention**: **Performance of the perceptual understanding** of a scientific image depends on the **epistemic uses of perceptual attention**, conceived of as the **system which controls perceptual and motor routines** to resolve pragmatic and epistemic queries:
(H1) in order to evaluate **perceptual predicates** (or observational propositions) about the elements presented by the image's surface and to diagnose the presence of recognizable or analyzable contents; and
(H2) in order, ultimately, to **access information about the referents** of the image via the singular knowledge of elements displayed by the image's surface.

4.2 'Attention' in *H*

Traditionally studied in psychology in which it has been frequently conceived as the **limited** faculty/capacity of **selecting information for further processing**, e.g. in the visual field.

Here: a **system of control of sensory-motor routines or skills**.

Procedural theory of attention: account that view attentional capacities as **being coincident with the exercise of epistemic and pragmatic procedures that are dependent on a context of use and of strategies for reaching particular goals**. According to this type of analysis, attention uses strategic and exploratory operations that enable the agent to obtain information (typically) on a particular target element or cues related to one object or spatio-temporal element (and often also to constitute a singular representation of this target).

4.3

One can give an account of this strategic structure by analyzing selective attention as being dependant on two main components: (i) a set of **instructions for the control of bodily or mental events**, which can be termed either epistemic queries or pragmatic queries – and (ii) a set of **elementary operations called routines** that allow, according to varied and context-dependent combinations, to give an answer to or to satisfy the epistemic and pragmatic queries. The concept of routine refers to the perceptual or motor elementary procedures that can be used to satisfy or solve the queries (epistemic or pragmatic) on the basis of the evaluation of perceptual predicates. Attention, as a mediating faculty, seems to be the capacity that organizes the relations between conceptual and non-conceptual routines for demonstrative identification.

5.2 Specific argument from combinatorial arrangement of elements

About the 'navigation' in the **compositional structure** of the image: **only the attentional system allows perceivers to retrieve the compositional structure of the image**, that is, the spatial relations among the elements displayed by the image's surface.

The support for this idea can be derived from **the need to appeal to visual routines to explain the visual understanding of basic spatial relations**, such as in Ullman's [8] analysis.

Let us assume a distinction between the 'automatic' (or 'stimulus-driven') formation of 'early visual representation' and sub-sequent application of visual routines. The argument is that a number of operations performed during the examination of an image require performing visual operations which, arguably, cannot be accounted for neither by so-called 'automatic' or 'stimulus-driven' processing nor by 'purely conceptual' abilities based on type/kind identification. One can argue that the control of visual routines is required namely for operations such as retrieving basic elements or shapes in a display and specifying the spatial relations among these basic elements/shapes (top/bottom, right/left, inside/outside etc). Ullman gives, for instance, the example of determining whether a point lies inside or outside a closed curve, and shows that a base representation of the point and the curve is not sufficient to resolve this query about spatial organization.

5.3 Specific argument from singular reference

Singular reference: reference to a token entity, e.g. a single physical object. If the perceiver has some mastery of the compositional characteristics of an image, this means that she **has been able to pick up elements** and examine their spatial relations. Given that a set of elements has been segmented, **a singular knowledge of the properties of each element becomes possible**, and this singular knowledge can serve as a background for the knowledge referential disposition of the image's parts. Building incrementally knowledge about the referent of an element *e* of an image (whatever this element might be) is prior to reasoning about the properties of the referent of *e*.

Building incremental knowledge about *e* is dependent on epistemic visual attention because such a framework assumes that that the singular knowledge of a particular element *e* is acquired incrementally by using sensory-motor routines – or visual object files – to evaluate perceptual predicates.

6.1 Example; photographs

The epistemic uses of photographs: each photograph carries information about the photographic referents (the objects that have been photographed). Cf. Figure 1.

The referential disposition of the photographic image is determined by a reliable causal process of image production. The camera has recorded, via a system using chemical or digital transduction, an optical projection of the referent. As a result, a number of properties of element *e* within the image (spatial, chromatic, textural properties) are counterfactually dependent on the properties of the referent *r* in a way that can be observed in a final photographic image.

The understanding of an epistemic use of a photograph (for example within a scientific discourse or practice) requires perceptual and conceptual abilities tied by epistemic attention – this is illustrated in the top triangle in Figure 1.

6.1.1

Entertaining **demonstrative thoughts** about elements *ei* of the photograph's surface – such as 'These dots are F' or 'This mark is an F' – **presupposes the perceptual segmentation of these relevant elements**.

For it is only once a primal segmentation of the elements is secured that the perceiver can formulate demonstrative identifications such as 'That element *e1* is an image of a shadow' etc.
Such demonstrative identifications of the elements allow the perceiver to form beliefs about relevant elements *ei* and their relations. For instance, on the basis of the perceptual identification of *e1*, the perceiver can evaluate perceptual predicates about *e1* such as *White*(*e1*) or *Touching*(*e1, e2*) or *Connected*(*e1, e4*). In addition, the perceiver can form singular beliefs about *e1* and its relations with other elements such as *e2* or *e3*.
(Generalization (considering e as a type) is dependent on demonstrative thoughts.)

6.1.2

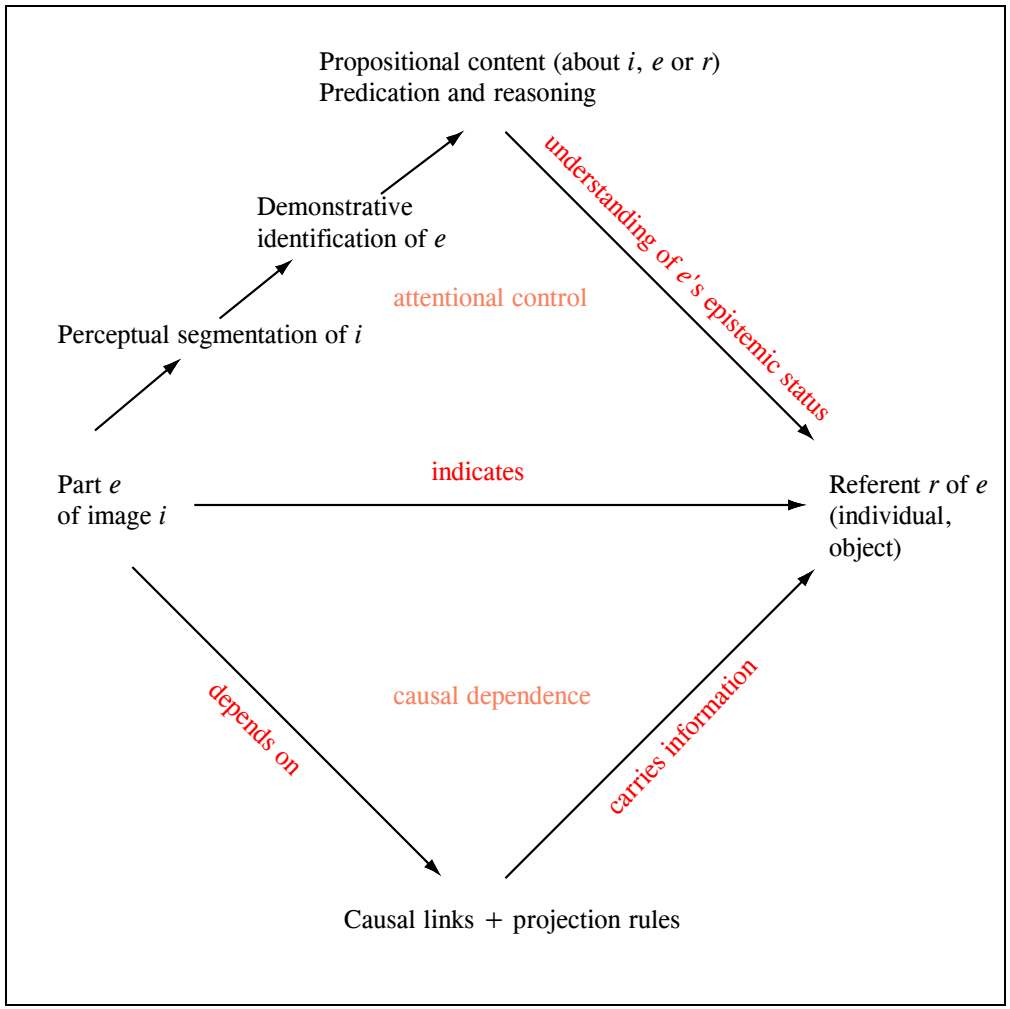

Fig. 1 Schema of a reference-based account of the epistemic uses of a photograph

6.2 Data from video

Experimental paradigms in psychology that use video recording and recorded images as data: For instance, Land, Mennie, & Rusted [19] recorded simultaneously (i) the activities of subject while performing the task of preparing tea and (ii) their eye movements. The aim of this study was to determine the pattern of fixations during the performance of a well-learned task in a natural setting (making tea), and to classify the types of monitoring action that the eyes perform. The authors used a head-mounted eye-movement video camera, which provided a continuous view of the scene ahead, with a dot indicating foveal direction with an accuracy of about 1 deg. A second video camera recorded the subject's activities from across the room. The videos have been linked and analyzed frame by frame. At least in this task, foveal direction was always close to the object being manipulated, and very few fixations were irrelevant to the task. According to the authors' classification, roughly a third of all fixations on objects could be definitely identified with one of four monitoring functions: (1) locating objects used later in the process, (2) directing the hand or object in the hand to a new location, (3) guiding the approach of one object to another (e.g., kettle and lid), and (4) checking the state of some variable (e.g., water level). If one question how scientists were able to reach such kind of interpretation for the collected data (large number of single video frames), it seems clear that **they have had to perform the tasks described by H**: segmenting, demonstrative identifications of the basic elements, and acquisition of singular knowledge about the referents of these basic element. Ultimately, **the classification of the four kind of monitoring functions of eye movements is based on assumptions about the referents of the elements of the image and their relations**: i.e., the eyes of the agent and the objects which are the targets of her or his actions.

7. Graph

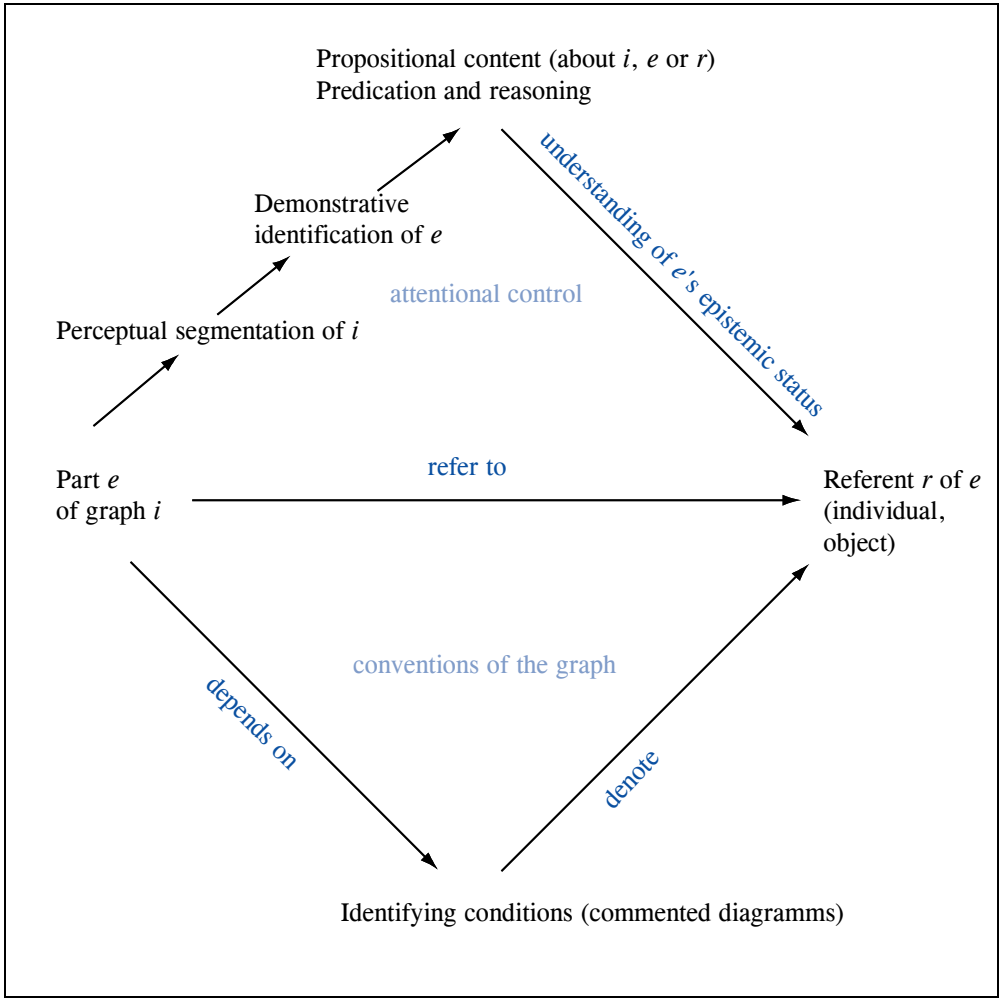

Fig. 1 Schema of a reference-based account of the epistemic uses of a graph

8. Questions for further research

Applied research in professional depiction systems of the CS of attention: How can CS of attention help in producing improved depictions for scientific research?

History and cognition of scientific images: How can we account for the relationships between the history of depiction systems and the perceptual faculties?

Tokens and types: What are the respective functions of the reference to types and the reference to tokens in scientific depictions?

Virtual reality: How does demonstrative reference work in the depictions used in virtual reality?

References

1. Goodman, N., The Languages of Art. 1968, Oxford: Oxford University Press.
2. Lopes, D.M.M., Understanding Pictures. 1996, Oxford: Oxford University Press.
3. Stenning, K., Seeing Reason, Image and Language in Learning to Think. 2002, Oxford: Oxford University Press.
4. Evans, G., The Varieties of Reference. 1982, Oxford: Oxford University Press.
5. Campbell, J., Reference and Consciousness. Oxford Cognitive Science Se-ries. 2002, Oxford: Clarendon Press.
6. Perry, J., Reference and Reflexivity. 2001, Stanford: CSLI Publications.
7. Miller, G.A. and P.N. Johnson-Laird, Language and Perception. 1976, Cambridge, MA: Harvard University Press.
8. Ullman, S., Visual routines. Cognition, 1984. 18: p. 97-159.
9. Pylyshyn, Z.W., Seeing and Visualizing: It's Not What You Think. 2003, Cambridge, MA: MIT Press.
10. Wright, R.D., ed. Visual Attention. 1998, Oxford University Press: New York, Oxford.
11. Pashler, H.E., The Psychology of Attention. 1998, Cambridge, MA: MIT Press.
12. Allport, A., Attention and control: have we been asking the wrong questions. A critical review of the last twenty five years, in Attention and Performance XIV, D.E. Meyer and S. Kornblum, Editors. 1993, MIT Press: Cambridge, MA. p. 183-218.
13. Allport, A., E.A. Styles, and S. Hsieh, Shifting intentional set: Exploring the dynamic control of tasks, in Attention and Performance XV: Conscious and Nonconscious Processing, C. Ulmità and M. Moscovitch, Editors. 1994, MIT Press: Cambridge, MA. p. 421-452.
14. Ullman, S., High Level Vision. 1996, Cambridge, MA: MIT Press.
15. Ballard, D.H., et al., Deictic codes for the embodiment of cognition. Behav-ioral and Brain Sciences, 1997. 20(4): p. 723-767.
16. Land, M.F. and S. Furneaux, The knowledge base of the oculomotor system. Philosophical Transactions: Biological Sciences, 1997. 352(1358).
17. Land, M.F. and M.M. Hayhoe, In what ways do eye movements contribute to everyday activities? Vision Research, 2001. 41: p. 3559-3565.
18. Kahneman, D., A. Treisman, and B.J. Gibbs, The reviewing of object files: Object-specific integration of information. Cognitive Psychology, 1992. 24(2): p. 175-219.
19. Land, M.F., N. Mennie, and J. Rusted, The role of vision and eye movements in the control of activities of daily living. Perception, 1999. 28: p. 1311-1328.