

perception

Oxford University Press

Oxford New York

Athens Auckland Bangkok Bombay

Calcutta Cape Town Dar es Salaam Delhi

Florence Hong Kong Istanbul Karachi

Kuala Lumpur Madras Madrid Melbourne

Mexico City Nairobi Paris Singapore

Taipei Tokyo Toronto

and associated companies in

Berlin Ibadan

Copyright (c) 1996 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.,

198 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Perception / edited by Kathleen Akins.

p. cm. — (Vancouver studies in cognitive science : ISSN v. 5)

Chiefly papers presented at a Vancouver studies in cognitive science conference, Vancouver, Canada, February 1992.

Includes bibliographical references (p. ).

ISBN 0-19-508461-6; ISBN 0-19-508462-4 (pbk.)

I. Visual perception—Congresses. 2. Perception—Congresses. I. Akins, Kathleen.

II. Series.

BF241.P397 1996

152.14—dc20 95-11810

## Acknowledgments

The majority of the papers in this collection were presented at a Vancouver Studies in Cognitive Science conference in February 1992 in Vancouver, Canada. This conference was funded by a grant from The Social Sciences and Humanities Research Council of Canada. I would like to thank the participants of that conference and the many members of the Simon Fraser University community who helped with its organization and financing: in particular, the VSCS committee members; Deans Robert Brown and Evan Alderson, the past and current Deans of Arts respectively; John Perry, director of the Cognitive Studies Programme; and Tanya Beaulieu, for organizational assistance. My gratitude as well to Eleanor O'Donnell, who copy-edited the volume, to Lindsey Thomas Martin, who prepared a camera-ready copy of the manuscript and to the Simon Fraser University Publications committee for their financial support. And finally a special thanks to Steven Davis, the general editor of this series, without whom neither the conference nor this volume would have been possible.

9 8 7 6 5 4 3 2 1

Printed in the United States of America  
on acid-free paper

## *Contributors*

- Kathleen A. Akins, Department of Philosophy,  
Simon Fraser University
- Dana Ballard, Department of Computer Science,  
University of Rochester.
- Patricia S. Churchland, Department of Philosophy,  
University of California, San Diego.
- Paul M. Churchland, Department of Philosophy,  
University of California, San Diego.
- Daniel C. Dennett, Center for Cognitive Studies,  
Tufts University.
- Frances Egan, Department of Philosophy,  
State University of New Jersey, Rutgers.
- C. Randy Gallistel, Department of Psychology,  
University of California, Los Angeles.
- John Grimes, Beckman Institute for  
Advanced Science and Technology,  
University of Illinois, Champaign-Urbana.
- John Haugeland, Department of Philosophy,  
University of Pittsburgh.
- John M. Henderson, Department of Psychology,  
Michigan State University.
- Kirk Ludwig, Department of Philosophy,  
University of Florida.
- Brian P. McLaughlin, Department of Philosophy,  
State University of New Jersey, Rutgers.
- Sarah Patterson, Department of Philosophy,  
Tufts University.
- V.S. Ramachandran, Department of Psychology,  
University of California, San Diego.
- Steven Winger, Department of Philosophy,  
University of Illinois, Champaign-Urbana.

# Contents

## Acknowledgements

### Contributors

1. **Introduction**, 3  
*Kathleen A. Akins*
2. **Explaining Why Things Look the Way They Do**, 18  
*Kirk Ludwig*
3. **A Feedforward Network for Fast Stereo Vision with Movable Fusion Plane**, 61  
*Paul M. Churchland*
4. **On the Failure to Detect Changes in Scenes across Saccades**, 89  
*John Grimes*
5. **On the Function of Visual Representation**, 111  
*Dana Ballard*
6. **Filling In: Why Dennett Is Wrong**, 132  
*P.S. Churchland and V.S. Ramachandran*
7. **Seeing Is Believing – Or Is It?**, 158  
*Daniel C. Dennett*
8. **Ships in the Night: Churchland and Ramachandran on Dennett's Theory of Consciousness**, 173  
*Kathleen A. Akins and Steven Winger*
9. **Lewis on What Distinguishes Perception from Hallucination**, 198  
*Brian P. McLaughlin*
10. **Intentionality and the Theory of Vision**, 232  
*Frances Egan*
11. **Success-Oriented and Individualism in Marr's Theory of Vision**, 248  
*Sarah Patterson*
12. **Objective Perception**, 268  
*John Haugeland*
13. **Visual Attention and the Attention-Action Interface**, 290  
*John M. Henderson*
14. **The Perception of Time**, 317  
*C. Randy Gallistel*

# 1

## *Introduction*

Kathleen A. Akins

This volume of the Vancouver Studies in Cognitive Science series arose out of the 1992 annual conference which was entitled "Problems in Perception." The point of such a broad topic, simply put, was to see what was new and interesting in the theory of perception across a broad range of disciplines – philosophy, psychology, computer science and the neurosciences. Despite the wide scope of the intended topic, however, there has been a fortuitous convergence in the subject matter of both the papers presented at the conference and those solicited for this volume: all focus on fundamental questions about the nature of visual perception. Taken together they ask "what is the nature of vision – what does vision *do*?" and more specifically "what are the form and content of visual representations – both of those representations involved in unconscious (or preconscious) visual processing and those that support visual phenomenology, our conscious visual perceptions of the world?"

I sit here at my dining-room table and look out at the panoramic view before me – the city skyline on the left, the snow-capped mountains in the background, the small float plane landing on the glinting water of the harbour. Despite the fact that the information received through human eyes is transient (e.g., it is interrupted by blinks and saccadic-eye movements), serial (I look, first, at one very small point in visual space then at another point) and varies in its informational "texture" (e.g., from the fovea to the periphery, the eye receives information with diminishing spatial resolution and frequency discrimination), the world we *perceive*, say, the panoramic view of which I am conscious, does not bear the mark of its representational origins. Even given the fragmented and serial nature of the incoming visual information, somehow we come to experience objects and properties as stable across time, as unified, as existing independently of our perceptions and actions, and so on. Focusing first on one small portion of the skyline and then on another, I nonetheless perceive the skyline as a whole, existing through time in front of me; the individual buildings do not seem to come into existence, one by one, as my eyes traverse the skyline

(first Harbour Centre, then Canada Place, then the Coliseum) nor are the individual buildings seen as fragmented or disunified. Moreover, in thinking about the visual experience itself, my perception of the skyline seems to be truly panoramic: at any moment, it *seems* to me that the visual field encompasses a single large "arc" of the world in front of me, one that is equally rich in detail across its entire broad field. In other words, we both perceive *the world* as independent, permanent and stable and we experience *visual phenomenology* as everywhere rich in information and "non-gappy." How, then, is this possible? How does the rich, unified panorama of human vision come about?

### Inference Theories and the Literalist View

Since David Marr, computationalists have assumed that vision is essentially the problem of determining "what is where." Given a series of retinal images (plus some fully general assumptions about the physical nature of the world) the task of vision is to produce, on the basis of a series of (unconscious) inferences, a veridical representation of both what is out there in the world (the properties of various objects and events) and where they are. For example, imagine the images that would be created on your retinas as, say, a golden retriever bounds towards you across the lawn. As the dog gets closer, the golden blobs in left and right images grow larger; across both images, the golden blobs begin to diverge – in the left image, the blob veers downwards and towards the left, while in the right image the blob moves down and to the right. From these two-dimensional images, the visual system must (among other tasks): discern figure from ground (pick out what, in the images, is the dog); "fuse" the two golden blobs of the left and right images (i.e., determine that there is but one golden dog in view); disambiguate relative motion from either expansion or rotation (i.e., determine that the dog is getting closer to you as opposed to remaining in place but expanding); decide that it is the dog that is moving and not yourself; determine the stable three-dimensional shape of the dog despite the changes in the dog's configuration caused by his bounding motion (i.e., his legs and back straighten and bend as he bounds, his ears flap up and down and, with some luck, his tail wags back and forth); and finally identify the dog as a dog (and indeed as your dog, Bowser). Thus, according to the standard computational account, you come to see the shaggy golden retriever, Bowser, bound across the lawn towards you: our veridical perceptions of events in the visual world are the result of a series of unconscious inferences, inferences based upon the information contained in the retinal image plus general background assumptions

about the world. This is what Kirk Ludwig terms an instance of an "inference theory" of perception.

Both Kirk Ludwig and Paul Churchland take issue, based upon the inferential nature of the processes posited, with this widely accepted view of the nature of vision. According to *Churchland*, inference theories – and hence standard current computational theories of vision such as Marr's – start off on the wrong foot by making two unwarranted initial assumptions. They assume, first, that our representations of the world (those that bring about our veridical perceptions) are basically *sentential* or propositional and, second, that the relationships between these pre-conscious states are of the sort "that standardly connect sentences together – relations such as entailment, coherence, probability, truth and so on" (p. 61). In his paper, Churchland argues for a different form of neural representation – high-dimensional activation vectors, or less formally, patterns of activation across neural populations – by presenting a vector activation model of stereo vision. Stereo vision is the ability to perceive the relative and absolute positions of objects based upon the relatively displaced images of the left and right eyes. While it is an intuitively simple task (when *we* look at two photographs taken from the vantage points of the left and right eyes, it is certainly easy for us to see that that is Bowser, just slightly displaced), providing a computational model of this ability has proved extraordinarily difficult – or rather the standard computational methods have failed to provide a model that will solve the problem in anything like real time. So Churchland's argument against the inference theory is a purely empirical one: he provides a vector activation model of a task, stereopsis, that standard computational models have (notoriously) failed to explain.

*Ludwig's* paper, on the other hand, offers an *a priori* argument against inference theories. The inferences postulated by standard computational accounts of vision, Ludwig points out, are neither conscious nor are they necessarily accessible to consciousness in principle – indeed it is precisely because such states are neither conscious nor necessarily accessible to consciousness that they are thought to explain the success of perception. But if this is so, Ludwig claims, then such inferences violate a "a central conceptual requirement" on the very attribution of mental states or processes to a person, namely they violate the "connection principle" (following Searle 1990). This is the principle that "nothing is a mental state unless it is a conscious mental state or it is a disposition to produce a conscious mental state." Ludwig argues that if such inferences cannot be attributed, qua mental states, to persons at all, obviously they cannot be used to explain or justify our veridical visual perceptions.

Daniel Dennett, Dana Ballard and John Grimes are also critical of the standard computational view but for rather more unintuitive reasons. In Ballard's terms, standard computational views – and this includes Churchland's own non-inferential but nonetheless computational theory of stereo vision – are committed to what Ballard calls the "Literalist View" or, to coin a phrase, the "You See What You Think You See" view. This is the commonsense position, the position we adopt, according to Ballard and Dennett, when we simply take our visual phenomenology at face value, when we try to model our theory of vision on how our visual experience *seems* to us, on the unreflective first-person perspective. To put this point in a slightly different way, (Dennett and Ballard) our models of vision start off on the wrong foot by endorsing a seemingly innocuous view: the belief that the *purpose* of vision is to produce visual perceptions – that the end product of visual processing is simply visual perception. Thus, most vision researchers (mistakenly) see their task as explaining *what we seem to see* – our conscious visual experiences – instead of investigating *what visual information is used to do*.

What exactly is the Literalist View, then, and how might it be wrong? After all, if one is having a veridical perception (e.g., the tree that seems to be before you *is* before you) what possible alternative could there be to "seeing what you think you see"? As Dennett and Ballard characterize the Literalist View, it appears to have at least four inter-related strands. First, as Ballard says, computationalists have assumed that visual phenomenology is (largely) the result of having "retinotopic representations in the brain that directly correspond to your conscious precepts." In its most naïve form, this is the view that the imagistic nature of visual perception results from "pictures in the head" – or, more generally (as Kathleen Akins and Steven Winger describe it), the commonsense view tries to explain the veridicality of perception by positing that properties of the world are represented by like properties in the brain, and that these representations, in turn, give rise to phenomenological experiences with similar characteristics. So, for example, a naïve person might imagine that viewing a red apple gives rise to an apple-shaped red representation in the brain, a representation which results in a visual image of a red apple, an image which is itself "red" and "apple-shaped" in some metaphorical sense. In its most sophisticated guise (i.e., in the form held implicitly by vision researchers), the Literalist View emerges as the assumption that the "imagistic" nature of visual phenomena is explained by spatial or "pictorial" or retinotopic neural representations (as opposed to, say, sentential ones).

Second, the Literalist View assumes that veridical perception involves (for lack of a better term) "complete" representation. For

example, when I look out over the Vancouver skyline and see it in all its glorious detail, this is surely because the detail I seem to see is represented, in its entirety, in a retinotopic or imagistic representation – or so the Literalist assumes. Both Artificial Intelligence (AI) and neurophysiological researchers, by adopting the Literalist view, often assume that the purpose of saccadic eye-movements (which serially fixate the retina on small portions of the visual scene) is to build up, bit by bit from these informational fragments, a complex representation of the scene – a representation that is as detailed as the visual experience *seems to be* to the viewer. By contrast, a non-Literalist approach ("You-Don't-See-What-You-Think-You-See") entertains the unintuitive possibility that visual representations, unlike visual *experiences*, may well be "partial" or "gappy" or "sketchy."

Third, and related, the Literalist assumes that such complete representations are built up as a matter of course, on-line, regardless of the current task(s) of the visual system (e.g., regardless of whether one is staring idly out the window, focusing a camera, or trying to identify a distant building). Again, a non-Literalist admits the possibility of what Ballard calls "just-in-time" representation or what Akins and Winger call "need-to-know access." If there is no pressing, imminent need for a detailed representation (say, no question "what building is that?" awaiting a reply), then no such representation need be constructed.

Fourth and finally, the Literalist posits that visual representations have a certain kind of content, a kind of content that goes hand in glove with the initial assumption of pictorial representation. A (distinct) picture of a brown speckled hen contains the information about every speckle on the hen (or at least, about every speckle on the visible portion of the hen!). This is part of the picture's content. The sentence 'there is a brown speckled hen', on the other hand, merely represents the hen *as* speckled; it does not contain any information, either explicitly or implicitly, about the individual speckles, or their spatial relations, one to another. It is this former kind of pictorial content, call it "explicit" or "concrete" content that the Literalist assumes must underlie our visual experiences. Again, if we see (in good light and at close range) a brown speckled hen, then what we see is what we think we see – we have imagistic representations the content of which is necessarily about every individual speckle.

A more general way of understanding the Literalist View is to see it as a compilation of two sets of common (inter-twined) intuitions. First, there are our assumptions about the relationship of the phenomenology of vision to the material events/neural representations that underlie them – about what kind of representations could possibly give rise to the phenomenology which we in fact experience. These intuitions, I suspect, are what motivate the Literalist assumptions that

visual representations must use retinotopic maps or images and that, as images, they have explicit or concrete content. Second, the Literalist View embodies the commonsense picture of *veridical* visual perception. We all assume that when we look out at the world, we perceive the world *as it really is* – veridically. All things being equal – assuming the viewer is not wearing distorting lenses or is swimming underwater or has just drunk a pint of scotch – what we see is what is actually there. But what kind of representations are required in order to see the world “as it really is”? What properties must the representations themselves have, what conditions must the perceptions satisfy? Here, veridicality, “seeing the world as it really is” seems to require (intuitively) that spatial properties of the world receive spatial representations (hence be pictorial), that they be complete and that have concrete content about the details of the visual scene. In other words, we can read the Literalist View as providing an intuitive set of necessary conditions for veridical vision.

As an antidote to the Literalist View, *Grimes* presents some fascinating experimental evidence that, during the normal inspection of a presented picture, we retain very little information – either because we fail to integrate the information across retinal fixations or because we simply do not process or retain very much of the available information at all. *Grimes* also suggests that integration, such as it is, may be more akin to abstraction than actual integration: when we retain the information about features of a scene, we first select the information on the basis of salience and then represent that information only abstractly – “a large hen, with many small speckles.”

*Ballard*, on the other hand, argues for what he calls the “functionalist view” (“You Don’t See What You Think You See”), the principal tenet of which is that “the machinery of the brain has to be accountable to the observed external behaviour” (p. 118). He therefore suggests that models of vision posit a “hierarchy of abstraction,” one level of which is the “Embodiment Level” – one that emerges between the neural and the cognitive levels – and that specifies how the facts of human embodiment constrain aspects of visual processing. More specifically, he presents a number of models for visual tasks that go against the Literalist paradigm by using “just-in-time” informational access to simplify the visual task and reduce memory requirements, object-centred (as opposed to viewer-centred or retinotopic) representations, “coarse-grained” or “abstract” information derived from non-retinotopic representations (“all the dots moving to the left”), the implicit representation of time and a “deictic” strategy for action.

In the papers contributed by *Daniel Dennett*, *Patricia Churchland* and *V. S. Ramachandran*, and *Kathleen Akins* and *Steven Winger*, the specific

issue at hand is the retinal blind spot. Because the axons of the ganglion cells merge together at one spot as they exit the retina, there is a spot on the retina that lacks photoreceptors; hence the retina has a spot, larger than the fovea, that simply fails to “see” a good portion of the world. Under normal viewing conditions – with both eyes open – the retinal blind spots are not noticed, for, in each eye, they “look at” slightly different parts of the world (thus the left eye sees the part of the world that the blind spot of the right eye misses, and vice versa). If only one eye is open, objects that fall entirely within the blind spot are not noticed. But if an object *spans* the blind spot, it appears to be “filled in.” *The gap or absence is simply not noticed. Why, and how, does this occur?*

*Churchland* and *Ramachandran* argue for a view of filling in that seems to fall more or less within the bounds of the Literalist camp. For example, *Churchland* and *Ramachandran* pose the nature of the debate in the following way:

What is going on when one’s blind spot is seen as filled in...? Is it analogous to acquiring a nonvisual representation (belief) that Bowser, the family dog, is under the bed, on the basis of one’s visual perception of his tail sticking out? Or is it more akin to regular visual perception of the whole Bowser in one’s peripheral but non-blind field? That is, is the representation itself a visual representation, involving visual experiences? (P. 133)

Here, as later passages make clear, *Churchland* and *Ramachandran* assume that there is a distinction between inferential conclusions/sentential representations and “visual” ones, that visual representations can be equated with the known retinotopic maps/representations of various visual areas (such as the retinotopic map of V1) and that there is a corresponding distinction to be drawn between the contents of visual representations and inferential conclusions (one simply *sees* the tail directly, whereas one *infers* that the dog is under bed). They argue on the basis of empirical evidence that, *contra* *Dennett*, filling in (at least in some cases) is the result of interpolation mechanisms that serve to complete complex representations at the level of visual or retinotopic representations.

*Dennett’s* response to *Churchland* and *Ramachandran* takes issue with the distinction (in his terms) between the perceptual and the conceptual – a distinction that he sees as part and parcel of the “Myth of the Cartesian Theatre,” the view that there is a place in the brain where “it all comes together,” where consciousness occurs. By *Dennett’s* lights, to posit that visual processing eventuates in a conscious perception (seeing the tail of Bowser poking out from under the bed)



and that, on the basis of such conscious events, we then draw conclusions about the further nature of the world, unseen but inferred (that Bowser is under the bed), is to accept a fictitious divide – a juncture before which visual representations are unconscious and after which, in the form of perceptions and conceptual conclusions, consciousness occurs. In other words, Dennett ties the Literalist View to the Myth of the Cartesian Theatre and then argues against Churchland and Ramachandran (and thereby against the Literalist view of filling in) largely on general considerations about the nature of consciousness.

In the article by *Akins* and *Winger*, an attempt is made to sort through what, precisely, divides the two camps. As the title suggests (“Ships in the Night”), *Akins* and *Winger* see the two parties as, to a large extent, talking at cross-purposes. Although often couched as an empirical dispute, at bottom, what divides Dennett from Churchland and Ramachandran are fundamental philosophical questions about the relationship of mental representation to conscious experience.

Finally, in the paper by *Brian McLaughlin*, the very general question of the veridicality of visual perception is addressed, here in the guise of the traditional philosophical question about the difference between veridical perception and hallucination. *McLaughlin* addresses David Lewis’s counterfactual theory of vision, a view that, when stated in its most intuitive form, is very much in the Literalist tradition. As Lewis tells us, “it is not far wrong to say simply that someone sees if and only if the scene before his eyes causes matching visual experience” (p. 1986). (Here, a visual experience can be said to “match” an environmental scene to just the extent that the scene satisfies the content of the experience, while “the content of the experience” is defined by *McLaughlin* as “roughly, the way things look to the subject of the experience in virtue of her having the experience.”) The problem, however, is that there are many counterexamples to the intuitive claim, in particular cases of “veridical hallucination”. In such cases, the subject hallucinates a particular scene and, as a matter of pure coincidence (trickery, magic or what have you), the hallucinatory experience matches the scene before the subject’s eyes. (E.g., With my eyes closed, I hallucinate that a large black bug is crawling across my desk just as, as a matter of fact, a large black bug does crawl across my desk.) In order to preclude such counterexamples, Lewis adds a counterfactual condition to the original conception. He says:

This is my proposal: if the scene before the eyes causes matching visual experience as part of a suitable pattern of counterfactual dependence, then the subject sees; if the scene before the eyes causes matching visual experience without a suitable pattern of counterfactual dependence, then the subject does not see. (1986, 281)

*Lewis* further spells out this condition in more detail, settling upon a formulation which he calls the “multiple tracking condition,” a definition that excludes any explicit reference to causal relations between the scene and the experience. *McLaughlin*’s response to *Lewis* is that the multi-tracking condition is at best a criterion for when the capacity to see is being exercised, not an explanation of the capacity itself. That is, *McLaughlin* suspects that *Lewis* has simply reversed the order of explanation: he has mistaken the conditions that hold *as a result of* the capacity to see for a substantive explanation of the nature of that capacity, of the nature of veridical vision. But if this is so, why then does the multi-track condition “work”? Why does it, by and large, exclude cases of visual hallucination? As *McLaughlin* says, a wide pattern of counterfactual dependence provides us with good (although defeasible) evidence that a particular kind of mechanism is at work, a mechanism that serves to discriminate between visual scenes of various kinds. *McLaughlin* ends his article, then, with a rough proposal that equates the capacity to see with the capacity to make certain kinds of visual discriminations. (Note here that *McLaughlin*’s counterproposal is also a Literalist View, for it takes the content of visual experiences at face value and makes us of the undischarged notion of “matching” content.)

### The Intentional Content, Objectivity and Unity of Visual Perception

In the articles by *Frances Egan*, *Sarah Patterson* and *John Haugeland*, it is the intentional content of visual representations and perceptions that is at issue. Both the *Egan* and *Patterson* papers are explicitly engaged in the contemporary philosophical debate over “externalism” – roughly put, the view that mental/neural events are individuated by means of their content or meaning, and that the content of psychological events is itself individuated by events, entities or properties external to the subject. To put the debate in context, *Jerry Fodor* (1980) has claimed that the very possibility of computational theories of mind – and of cognitive science in general – depends upon the individuation of mental/computational states solely on the basis of properties *internal* to the subject, or as he would say, on the “formal” or “syntactic” properties of computational states. In response, *Tyler Burge* (1986) has argued that, in fact, typical computational theories of perception, such as *Marr*’s theory of vision, actually presuppose an externalist account of content individuation and, hence, of mental state individuation as well. Under normal conditions, *Burge* claims, given the evolutionary adaptation of our sensory systems to our present environment, it is assumed that our sensory systems are “successful,” that we perceive what is actually there – in other words, that our perceptions are veridical. (Referring to

Marr's theory of vision, Burge calls this the "success-orientation" of Marr's theory.) Moreover, it is the normal distal causes of sensory representations that define their intentional contents: "the information carried by representations – their intentional content – is individuated in terms of the specific distal causal antecedents in the physical world that the information is about and that the representations normally apply to" (1986, 32). Thus, for example, the dotted-lines in Marr's 2½-D sketch represent discontinuities in object surfaces – are "edge detectors" – precisely because the algorithms that produce dotted lines from information contained in the Primal Sketch are designed to produce these primitives when and only when an "edge" is present. (In the case of our own visual systems, of course, the notion of conscious design is replaced by that of evolutionary adaptation within a specific environment; if there are representations of edges in our visual system, it is just because there have come to be regular, law-like relations between certain syntactic elements and edges of objects in the world.)

In this volume, both Egan and Patterson take issue with Burge's understanding of cognitive psychology in general, and in particular, his externalist interpretation of Marr's theory of vision. Egan's main argument hinges upon the formal nature of Marr's theory. Computational theories, Egan argues, are formal, mathematical constructs and as such they carry with them no *essential* interpretations. This is not to deny that interpretations based on external causal factors are commonly and legitimately made. Given, say, a raw primal sketch derived from an actual photographic array, the primitives of the sketch are reliably correlated with intensity changes across the retinal array. Hence they can be legitimately interpreted as representing intensity changes. Moreover (as Burge would agree) representational primitives "have no content considered independently of the particular environment in which the subject is normally situated." Had the photographic array been produced in a world with quite different light properties, the representational content of those primitives would likewise be different. On the other hand, *the facts of evolution* – of the adaptation of the organism to a specific environment – *do not serve to fix any interpretation as essential to the individuation to the primitives.* Such facts could only give us *good reason* to see the interpretation as privileged, for given the formal nature of the computation, no interpretation can be considered essential. Rather, ascribing intentional content plays three pragmatic roles for the researcher: first, the intentional account makes perspicuous the formal account by stating, informally, the nature of the function performed (e.g., the module is computing depth); second, the intentional interpretation of representational primitives and processes allows us to see how, overall, the

problem of vision is solved – it allows us to see how the formal structures of the model solve the problem at hand; and third, an intentional interpretation allows us to see how the separate modules of vision are tied together, the role that each module plays in the overall functioning of the system.

What, then, of some other non-external means of individuating the intentional content of computational states? That is, what of the proposal that the primitives of Marr's theory of vision are indeed intentional, but that their content does not depend upon external causal factors? Egan ends her article with an argument against such "narrow content" views. She argues that even if one could square such an account with Marr's intentions, and even if one could actually construct the right sorts of contents based upon only "internal" considerations alone, such narrow contents would not yield any additional explanatory power over and above that provided by a formal characterization of the computational process.

In Patterson's response to Burge, she questions Burge's interpretation of Marr's assumption of success, arguing that this assumption cannot be equated with the presupposition of veridical perception. Rather, the assumption of success is best explained as a *methodological dictate* for artificial intelligence. If we wish to understand the ways in which the human mind works, our best chance of doing so lies in making an intelligent choice about the kinds of human cognitive abilities to be modelled. We should choose tasks that people perform both well and regularly, such that we can formulate a concrete characterization of the tasks performed. Thus in setting the computational task, we must look to the human case – to see what we do, what we do well and, in the unusual circumstance, how we fail when things go awry. Marr's assumption of success, according to Patterson, is merely the dictate that we model what people do well – and such an assumption is a long way from the assumption of veridical perception. Hence, Burge cannot assume that the individuation of visual representations will necessarily depend upon factors external to the individual, that in counterfactual situations, both content and individuation will turn upon "normal causes."

In contrast to Egan and Patterson, Haugeland's view of intentional content is what one might call "radically externalist" – or at least this might be said of his account of one class of perceptions, "objective" perceptions. Objective perceptions, on Haugeland's view, involve the perception of *objects qua objects*. Thus I see the distant bridge over the harbour *as a bridge* – as an object of a certain type, as a cohesive spatio-temporal unit that instantiates a set of mutually compatible and interdependent properties. The question of objective perception, however,

is not merely the question of how it is that, say, my visual perception comes to be about *that thing*, the bridge in the distance, for this way of phrasing the problem already grants objects qua objects a pre-theoretic (metaphysical) legitimacy. Rather, Haugeland claims, "the 'objecthood' of perceptual objects and the 'of-ness' of perceptions go hand in hand, and are intelligible only in terms of one another.... So the deeper question is: *How and why* is such a structure – what one might call the structure of objectivity – imposed on the physics and physiology of sensation?" (p. 273). Haugeland's answer turns on the constitutive standards that govern the "unity" or the "integrity" of objects qua objects. For example, any object, at a given time, if it is an object at all, must exist in one and only one place, have a particular shape, size and so on. Further, its properties must be mutually compatible: given a certain property *p* (say, the property of being spherical), there may be other properties, *q*, *r* and *s* that the object must have as well (say, having a certain size, certain surface texture) and other properties, *x*, *y* and *z*, that it lacks or could not have (it cannot also be a cube or a pyramid or shaped like a donut). Moreover, these constitutive standards, like those of chess and other games, are ineliminably *social*. In order to perceive objects as objects – to have objective perceptions – we must be able to understand the social standards to which such perceptions are (necessarily) held.

Finally, in the last two papers, contributed by *Randy Gallistel* and *John Hendersen*, the issue is also the unity of our visual perceptions, although here under the guise of what is commonly known as the "Binding Problem." The Binding Problem (or, as there are a number of different binding problems, the problem as portrayed by Hendersen and Gallistel) is roughly as follows. Given a single object in the world, the visual system represents its various properties at a variety of spatially distinct locations in the brain – in a number of different modules each of which has, presumably, its own distinctive system of representation. So the representations of the object are, in some sense, a spatially disjoint and representationally disunified set. On the other hand, our perceptions of that object are not perceptions of a set of disunified or disorderly properties at all. As Haugeland says, objects have integrity – they are, qua objects, unified wholes.

*Hendersen's* view of unification follows from his more specific hypothesis about the nature of visual attention, the hypothesis that there exists a "rubber band" relationship between visual attention and eye-movement. After the currently fixated location is processed, Hendersen contends, attention shifts to a new location. This location is determined, in the usual case, by pre-attentive, low-level feature analysis combined with a weighted choice structure for the various

features so determined. (At other times, higher-level processes can co-opt the attention mechanism, such as when, in reading, semantic considerations require a backwards eye-movement.) At the same time as attention is shifted, a signal encoding the new location is sent to the oculomotor system, while the allocation of attention to a new stimulus instigates higher level analysis of that stimulus. Then, after a lag caused by oculomotor programming, the eyes "snap" to the new position, to what is now the focus of visual attention.

How, then, does this model of visual attention solve the binding problem? Following *Triesman's* "Feature Integration Theory" (1986), *Hendersen* believes that individual properties of a visual scene are represented on modular feature maps. These maps are each linked to a "master map" of stimulus locations. When attention fixes on a particular place on the master map – in other words, when attention is fixed on a stimulus in a particular location – the various properties of that stimulus are serially accessed from the individual feature maps through their particular spatial locations on those maps. (In other words, the master map gives an explicit representation of space while modular feature maps only encode features *as* occupying a particular place, implicitly.) The features thus selected are then integrated to form a unitary representation that specifies their structural relations. Here, *Hendersen* postulates that visual attention, in addition to binding the visual properties of an object, may also combine visual and motor representations. There are two ways this might work. First, when visual attention has made explicit a spatial location of the new stimulus, its co-ordinates might become available to a variety of motor systems that are involved in programming movements directed towards that stimulus (e.g., those readying the brain to, say, grasp the object, look at it or lean towards it). Second, perhaps the explicit representation of the spatial location is used to decide among various competing contradictory motor signals within a single motor system (for example, there is some evidence that a variety of contradictory motor signals for saccadic eye-movements exist at one and the same time and hence that some selective process must be at work). Thus selecting a stimulus for further visual processing also serves to select one action among a host of conflicting ones. In either case, *Hendersen's* point is that part of what we know of an object as an object is its relationship to our own actions, in addition to our understanding of it as having a unified set of visual properties.

*Gallistel's* answer to the Binding Problem is what he calls a "recollective" theory of unified perception. Recall that on the *Triesman* theory of feature integration, attention unifies perception by accessing the features of a stimulus based upon a location on the master map.

Triesman also suggests that primitive object markers or “object files” are created. Each such file contains only a record of the spatial position (given in the coordinate system of the master map) of the object across time – a space-time path of the particular object. It is to these files that the various perceptual systems add information about the other properties of the object – about its shape, colour and so on. In contrast, Gallistel claims that “the records of experience in memory are as fragmented as the sensory perceptual process itself” (p. 333) In each sensory module, there are records with two distinct “fields” (or forms of information): the first field constitutes a space-time “worm,” a description of place as a function of time, while the second field specifies certain stimulus properties (given in the dimensions of that particular sensory space) as a function of time. Thus each sensory module keeps its own records of the position in space and time of various sensory properties. Unification begins with the retrieval of a sensory record based upon a particular salient property – say, the colour or taste of an object which is currently of interest to the organism. Based upon the salient property, the correct sensory files are found; the spatial records of that file are then used to probe various other sensory records – that is, to determine the other properties of the object. Thus, unification occurs through “recollection” – only when information about an object or property is required at some time after initial perceptual processing.

Note here that Gallistel’s general theory of mental representation exemplifies aspects of both the Literalist and non-Literalist views. Following the non-Literalist, Gallistel advocates a view of perception that does not take unified perceptual experience as the end state of perceptual processes. Rather unification occurs as a result of a particular behavioural need – the retrieval of unified sensory information occurs in order to carry out some behaviour task. On the other hand, Gallistel believes that each sensory representation in every sensory module is tagged with place and time markers regardless of the system’s present task or particular needs; moreover representations themselves are defined by Gallistel (1990) as being structures that are “isomorphic” with certain features of the environment. So these latter aspects of the theory better conform to the standard Literalist View.

### References

- Burge, T. (1986). Individualism and psychology. *Philosophical Review*, 95: 3-45
- Fodor, J. (1980). Methodological solipsism considered as a research strategy. *Behavioral and Brain Sciences* 3: 63-110
- Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press

- Lewis, D. (1986). Veridical hallucination and prothetic vision. In Lewis (1986) *Philosophical Papers: Vol. 2*. Oxford: Oxford University Press, 273-86
- Searle, J.R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences* 13: 585-96
- Triesman, A. (1986). Features and objects in visual processing. *Scientific American* 255 (5): 114-126