As an example of this approach, I have developed an oculomotor control model that simulates the eye movements of an infant as it tracks moving objects (Schlesinger & Barto 1999; Schlesinger & Parisi 2001). Using a reinforcement-learning algorithm, the eye-movement model quickly learns to anticipate the reappearance of an occluded object. Perhaps more impressively, like the young infants in Baillargeon's (1986) study, the model also tracks the target appropriately when its path is blocked by a hidden obstacle. Note that these prospective behaviors are achieved with only a simple associative learning mechanism, and without assuming or building in the capacity for memory or prediction.

Obviously, this eye-movement model is just one small step toward trying to explain infants' visual cognition in non-representational terms. Nevertheless, the model suggests that we can know about and act in the world, indeed, even anticipate future states of the world, without having to explicitly reference a mental model or symbolic representation. In this sense, of course, an important point of agreement between my model and **O&N**'s more general theoretical account is that *visual cognition is activity.* It is unclear to me, however, how far such a theory can go toward explaining not only the (apparent) capacity for infants to represent the world in symbolic form, but also their capacity for using those putative representations to act, judge, or reason about the world in a prospective manner.

Summary. I have highlighted here the representational account of infant cognition. My challenge to **O&N** is to continue to expand and elaborate their theory in at least two directions. First, I hope that they can flesh out the sensorimotor account of vision in enough detail to provide a compelling alternative to the idea that knowledge acquisition is simply a process of copying external reality. Second, and more generally, I encourage both authors to raise their sights, and to begin thinking beyond the question of learning and, instead, toward the lifelong process of perceptual and cognitive development.

ACKNOWLEDGMENTS

This commentary was supported by a grant from the National Institute of Child Health and Human Development (1R03 HD40780-02). I thank Michael Young for his helpful comments during its preparation.

Change blindness, Gibson, and the sensorimotor theory of vision

Brian J. Scholl[a] and Daniel J. Simons[b]

[a]*Department of Psychology, Yale University, New Haven, CT 06520-8205;*
[b]*Department of Psychology, Harvard University, Cambridge, MA 02138.*
Brian.Scholl@yale.edu dsimons@wjh.harvard.edu
http://pantheon.yale.edu/~bs265
http://www.wjh.harvard.edu/~viscog/lab/

Abstract: We suggest that the sensorimotor "theory" of vision is really an unstructured collection of separate ideas, and that much of the evidence cited in its favor at best supports only a subset of these ideas. As an example, we note that work on change blindness does not "vindicate" (or even speak to) much of the sensorimotor framework. Moreover, the ideas themselves are not always internally consistent. Finally, the proposed framework draws on ideas initially espoused by James Gibson, but does little to differentiate itself from those earlier views. For even part of this framework to become testable, it must specify which sources of evidence can support or contradict each of the component hypotheses.

On its surface, the "sensorimotor account of vision" by **O'Regan & Noë** (**O&N**) has an impressive scope: among many other things, it allegedly explains the nature of visual consciousness; how sensory modalities differ; how sensation differs from perception; how we perceive a stable world despite eye movements; why visual binding is unnecessary; and why hundreds of years of philosophical analysis of the problem of qualia can be dismissed as misguided.

How can so many phenomena fall under the explanatory scope of this single theory? One reason, we suggest, is that it is not so much a coherent theory as an unstructured collection of three interesting ideas: (1) vision is active and exploratory rather than passive; (2) knowledge of sensorimotor contingencies plays a central role in conscious vision; and (3) the visual system uses the world as an "outside memory." Although most researchers would accept the first idea, the latter two are more controversial. More importantly, these ideas are essentially unrelated: each can be selectively denied while maintaining the others.

The scope of these hypotheses and the many types of evidence alleged to support them creates a substantial problem: given that the theory is not a single, structured claim, it is unclear which of the ideas are supported by which types of evidence. The sensorimotor theory is treated by **O&N** as a coherent whole, and evidence consistent with some of the ideas is inappropriately taken to substantiate the theory in its entirety.

The role (or lack thereof) of change blindness in the sensorimotor theory. Here we focus on just one instance of this error, involving *change blindness* – the phenomenon wherein surprisingly large changes go unnoticed, even when observers are actively trying to find them (see Simons 2000a). The authors view change blindness as central to their theory: "the sensorimotor approach to vision . . . has provided the impetus for a series of surprising experiments on what has come to be known as change blindness. The robustness of these results in turn serves to vindicate the framework itself" (sect. 9). We suggest that both of these claims are mistaken, and that change blindness does not directly support the sensorimotor theory.

O&N suggest that change blindness was discovered as a direct consequence of the sensorimotor theory, or more precisely, the "world as an outside memory" claim. Although this idea did provide some of the theoretical motivation for recent work on such phenomena, the initial work on change blindness was not motivated by this issue at all. Most early work on change blindness derived from the study of visual integration and focused on the detection of changes during reading. For example, McConkie and Zola (1979) showed that observers often failed to notice when every letter on a screen changed case during a saccade. Other work on the failure to notice changes, both theoretical and empirical, similarly predated the current theory (e.g., Dennett 1991; Hochberg 1986; Phillips 1974; Stroud 1955).

The notion of using the world as an outside memory (e.g., Brooks 1991; O'Regan 1992; Stroud 1955) might explain why several forms of change blindness occur: we intuitively expect to detect such changes, perhaps on account of implicit beliefs about the capacity and fidelity of internal representations, or perhaps because of implicit expectations about the range of unusual or distinctive events that will draw our attention (e.g., Levin et al. 2000; Scholl et al., submitted). In any case, accurate change detection, when it does occur (in situations which do not induce change blindness) may be driven largely by motion transients which draw attention back to the world itself (e.g., Simons et al. 2000). Though the externalized memory hypothesis might predict change blindness, however, it is not clear that the sensorimotor hypothesis would. Sensorimotor contingencies require an internal memory from one instant to the next, because detecting contingencies depends on the ability to note how an environment changes in response to actions such as a "flick of the eyes." However, if the observer relied solely on the external world to provide their memory, then nothing would ever be seen to change across such flicks of the eyes (due to saccade-contingent change blindness). How, then, would observers learn what was stable and what was variable over time and across eye and head movements?

Thus, change blindness – including the flicker and mudsplash paradigms developed by Rensink, O'Regan, and colleagues – provides no direct support for **O&N**'s sensorimotor contingency idea. In fact, it does not even directly support the externalized memory idea. Change blindness is consistent with the idea that we lack internalized, detailed representations: in the absence of such inter-

nal representations, change blindness would occur. However, the existence of change blindness does not logically require the absence of a representation (see Simons 2000b). Representations are needed to detect change, but they could also be present in the face of change blindness. For example, observers might represent both the original and changed scene, but simply fail to compare them directly (Levin et al., in press; Scott-Brown et al. 2000; Simons 2000b; Simons et al., in press). The presence of change blindness allows no conclusive inferences about the presence or absence of internal representations. All it tells us is that if we have such representations, we do not spontaneously gain conscious access to the differences between them.

Even if we grant the idea of an externalized memory, however, we can simultaneously deny every other aspect of the theory – including the idea that knowledge of sensorimotor contingencies plays a substantive role in conscious vision. Change blindness provides little support for the externalized memory idea and provides even less for the framework as a whole. It is thus misleading to characterize change blindness as vindication for the overall framework. At best, one small (and dissociable) part of the theory is *consistent* with change blindness.

Are sensorimotor contingencies truly sensory? Do we really lack representations? The problems induced by the lack of integration among the central ideas in **O&N**'s framework are amplified by the fact that the individual ideas are not always internally consistent on their own. Take, for example, the central idea that perception derives from knowledge of sensorimotor contingencies. This claim depends on a consistent relationship between objects in the world and changes in retinal stimulation. Essentially, these retinal changes must reveal invariant properties of the objects. However, in describing this proposal, O&N seem to want the theory to include not just flicks of the eyes, but also flicks of attention. But in what respect is a shift of attention *sensory*? If it does not change the retinal stimulation, how can it be the basis of a sensorimotor contingency?

Similarly, it is not clear that this model truly lacks internal representations and memory. The repeated appeal to "knowledge" of sensorimotor contingencies seems little different from an internal memory or representation of an object. The only difference from a traditional object representation is that the "knowledge" in this case is of dynamic rather than static information. It was for this reason that other attempts to eradicate memory from the process of perception argued that the invariant information underlying perception was present externally, in the environment (e.g., Gibson et al. 1969).

Gibson redux? The notion that manipulation of sensorimotor contingencies underlies perception and awareness is old, dating at least to behaviorist views. To quote an example from that era: "The awakening of a retained sensory impression when its response is made is memory in the common sense of the word. Thought, then, appears as a means of 'trying' different actions and anticipating their results through a process of automatic recall" (Ross 1933). Although **O&N** acknowledge the prior related work of several other authors, they pay surprisingly little attention to the one researcher who most (in)famously proposed these ideas: James Gibson. Both the content of the sensorimotor theory as well as the style of its exposition are highly reminiscent of Gibson's work on direct perception. Yet, O&N rarely mention these similarities, even when Gibson addressed the same issues extensively. For example, both Gibson and O&N argue that perception is exploratory, that it depends on detecting the constant information amidst change (invariants), that learning improves the detection of correlations, that temporarily occluded objects are "seen" (Gibson et al. 1969), and that vision does not rely on an internal representation of the world (see Gibson 1966; 1979/1986 for further details). Furthermore, both essentially define away classic problems in perception and cognition such as the binding problem (O&N), or perceiving depth (e.g., Gibson 1966, p. 298). Despite the overwhelming similarity between the approaches and their implications, O&N entirely neglect discussion of how their views differ from Gibson's, or of the ways in which their approach might better handle the many well-known critiques of direct perception (e.g., Fodor & Pylyshyn 1981; Ullman 1980). Given the similarities, it is not clear that O&N fare any better in the face of such critiques.

One fundamental difference between Gibson and **O&N** is the nature of the stimulus for perception. Gibson emphasized that the information for perception was available in the visual world rather than in the retinal stimulation. Observers generally do not become aware of the retinal sensations produced by objects – they are just aware of the objects. Although O&N appeal to similar notions, their argument relies more on invariants of the retinal stimulation as the basis for perception. This issue is perhaps most apparent in the discussions of seeing "red" and of "filling in." In these sections, the authors focus on how the brain might perceive changes to the retinal stimulation that result from moving the eyes, noting that the blind spot might provide additional information due to the changes in stimulation it imposes.

This approach has its pitfalls. For example, the theory is difficult to test empirically because it is affected by the nature of the sensory apparatus to a greater degree than Gibson's views. One advantage of Gibson's approach is that evidence for the presence and use of environmental invariants in perception could be taken as support for the theory. For Gibson, invariant information is present in the environment regardless of whether or not the perceiver is capable of "picking up" that information (Gibson 1966). For O&N, the information available for perception depends critically on the nature and structure of the sensory apparatus (at least in some cases – for other forms of sensorimotor contingency, they seem to appeal to a more Gibsonian view). The particular sensorimotor invariants that define red for one observer might then not be identical to those that define red for another observer. Although their approach to defining "redness" is clever and original, it is also untestable because no *laws* of sensorimotor contingency can be specified; their invariants are tied to the sensory apparatus and will not generalize from one observer to another. Not surprisingly, then, their paper lacks details about the nature of the contingencies that could underlie the perception of "red."

Concluding thoughts. The sensorimotor theory of vision is notable largely for its impressive breadth: it attempts to marshal a wide variety of evidence in support of its several ideas, and thereby attempts to explain (or define away) several longstanding puzzles about the nature of visual experience. Although individually these ideas are each intriguing**, O&N** do little to explain how they are interrelated and how the framework as a whole is structured. Consequently, evidence that is relevant to only one of aspect of the theory is often adduced as support for the whole. We have highlighted one example of this: whereas O&N claim that their theory is vindicated by their discovery of change blindness, we have argued that change blindness is entirely unrelated to claims about sensorimotor contingencies. Moreover, it provides little support for the more relevant claims about external memory. The lack of internal consistency, both of the framework as a whole and of its component ideas, leads to a view that is intriguing, but difficult to test empirically. Of course, some of these same objections have been applied to the quite similar views presented in Gibson's theory of direct perception (e.g., Ullman 1980).

In sum, the sensorimotor framework would be greatly clarified by considering in detail (1) which parts of the theory receive direct support from which types of empirical evidence, (2) which parts of the theory must stand or fall together, and (3) which parts of the theory are substantive departures from earlier Gibsonian arguments.[1]

NOTE

1. Following these suggestions might in some ways work in **O&N**'s favor by highlighting, for those who are not favorably disposed to the theory, which ideas needn't "go down with the ship." For example, we were intrigued by the idea of using sensorimotor contingencies to explain the difference between the various sensory modalities, but as it stands it is not apparent from the text that you can accept this idea while denying most of the other major claims.

ACKNOWLEDGMENTS

BJS was supported by NIMH grant No. R03-MH63808-01. DJS was supported by NSF grant No. BCS-9905578 and by an Alfred P. Sloan Research Fellowship.

The absence of representations causes inconsistencies in visual perception

Jeroen B. J. Smeets and Eli Brenner

Vakgroep Fysiologie, Erasmus Universiteit Rotterdam, NL-3000 DR Rotterdam, The Netherlands. **smeets@fys.fgg.eur.nl**
brenner@fys.fgg.eur.nl **www.eur.nl/fgg/fys/people/smeets.htm**

Abstract: In their target article, O'Regan & Noë (O&N) give convincing arguments for there being no elaborate internal representation of the outside world. We show two more categories of empirical results that can easily be understood within the view that the world serves as an outside memory that is probed only when specific information is needed.

In line with the arguments in the target article, we consider vision to be tightly coupled to motor control. In order to catch a ball, one needs information about its size, weight, position, speed, and direction of motion. These attributes are important for different aspects of the action, so that they can be determined and processed independently within what has become known as separate visuomotor-channels (e.g., Jeannerod 1999).

Although determining visual attributes independently might be useful for controlling actions, this does not mean that the outcomes are independent, because the laws of physics and geometry relate many of these attributes. For instance, if an object moves at a certain speed, its position will change at a corresponding rate. An internal representation of the outside world would combine all available information to yield the most likely (and thus consistent) representation. This would of course reflect the physical and geometrical relationships within the outside world. The consequence of independent processing is that the relevant sources of information are combined separately for each attribute. Physically related attributes might thus be determined on the basis of different sources, within physiologically independent pathways. If all attributes are determined veridically, this independence remains unnoticed. It becomes evident when the processing of one attribute is erroneous, as is the case in visual illusions (Smeets & Brenner 2001). Two examples clarify this.

For intercepting a moving object one needs information about its speed to regulate the timing of one's action, and information about its (egocentric) position to direct one's action. Due to the noisiness of extraretinal information on eye orientation, the most accurate estimate of object speed will generally be one based on relative retinal information (Smeets & Brenner 1994). For determining an object's egocentric position, the use of extraretinal information cannot be avoided. And indeed, moving a visual background influences the perceived speed, without influencing the perceived position (Duncker illusion). In our view, each such attribute is processed independently to control a certain aspect of our actions. The Duncker illusion therefore affects the timing of ones action, without influencing its direction (Smeets & Brenner 1995).

A similar reasoning holds for grasping an object to pick it up. To move the digits to the object's surface, information about positions on that surface is needed (Smeets & Brenner 1999). To subsequently apply adequate forces to lift the object, a visual correlate of the object's weight is needed: that is, its size (Gordon et al. 1991). As with the previous example, these geometrically related aspects (positions and size) might very well be determined on the basis of different sources of information. The positions will again be determined using extraretinal information, whereas the object's size might be determined purely on the basis of retinal information. This explains why illusions of size affect the lifting force in grasping, but not the grip aperture (Brenner & Smeets 1996).

Independent processing of physically related attributes is not only evident in the visual control of action, but also in conscious perception. For instance, if one looks for a while at a waterfall, and subsequently fixates a tree at eye-level near that waterfall, the tree appears to move upward. The apparent position of the tree remains approximately at eye-level. Other examples of inconsistencies can be found in visual illusions, such as the Müller-Lyer illusion. This illusion influences the perceived size of the figure without affecting the perceived positions of the end-positions (Gillam & Chambers 1985). In analogy to the claim that we process only one fragment of the world at a time (sect. 4.2), this apparent inconsistency suggests that conscious perception involves processing only one attribute of that fragment at a time.

If one accepts that not all attributes are processed at a time, one can understand the flash-lag effect (e.g., Nijhawan 1994). This effect manifests itself when a subject is fixating a screen on which a target is moving continuously while another target flashes. If the subject is asked to indicate the position of the moving target at the time of the flash, he will misjudge this position in the direction of the target's motion. This has been interpreted as the result of motion extrapolation. However, this cannot be so because if the target unexpectedly reverses direction near the moment of the flash, the misjudgements are never beyond the actual trajectory of the moving target. It is more likely to be caused by different processing times for flashed and continuously presented stimuli (Whitney & Murakami 1998).

However, there is no reason to assume that flashes are processed more slowly than continuously visible stimuli. What then is the cause of this apparent difference in processing time? If not all attributes are processed continuously, the position of the moving target will have to be probed at some instant. This presumably takes time, and can start only after the flash has been detected. The moving target's position (or other attributes such as its colour and shape) will be probed too late. If this explanation is correct, the flash-lag effect should disappear if we change the experiment in a way that allows the position of the moving target to be probed at the time of the flash. A simple way to do so is to provide an additional cue for the time (or equivalently, the position of the moving target) at which the flash will occur. Indeed, the flash-lag effect is reduced markedly when this is done (Brenner & Smeets 2000).

Re-presenting the case for representation

Benjamin W. Tatler

School of Biological Sciences, University of Sussex, Brighton, BN1 9QG United Kingdom. **b.w.tatler@sussex.ac.uk**
http://www.biols.susx.ac.uk/resgroups/scn/vision

Abstract: O'Regan & Noë (O&N) present the most radical departure yet from traditional approaches to visual perception. However, internal representation cannot yet be abandoned. I will discuss: (1) recent evidence for very short-term pictorial representation of each fixation; (2) the possibility of abstract representation, largely unconsidered by the authors; and (3) that sensorimotor contingency theory requires internal visual retention and comparison.

O'Regan & Noë (**O&N**) extend the implications of recent change detection studies by arguing that not only is it unnecessary for the visual system to construct a point-by-point pictorial representation of the world across multiple fixations, but that no such information need be internalised on even the shortest of time scales. However, the reader should be cautious before abandoning all notion of representation and should first consider some of the implications of this model and other possible accounts.

Whilst it would be hard to argue that we build up a point-by-