Consciousness and Cognition xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Bridgemanian space constancy as a precursor to extended cognition

Michael J. Spivey*, Brandon J. Batzloff

Cognitive and Information Sciences, University of California, Merced, United States

ARTICLE INFO

Keywords:
Space constancy
Visual perception
Embodied cognition
Extended mind

ABSTRACT

A few decades ago, cognitive psychologists generally took for granted that the reason we perceive our visual environment as one contiguous stable whole (i.e., space constancy) is because we have an internal mental representation of the visual environment as one contiguous stable whole. They supposed that the non-contiguous visual images that are gathered during the brief fixations that intervene between pairs of saccadic eye movements (a few times every second) are somehow stitched together to construct this contiguous internal mental representation. Determining how exactly the brain does this proved to be a vexing puzzle for vision researchers. Bruce Bridgeman's research career is the story of how meticulous psychophysical experimentation, and a genius theoretical insight, eventually solved this puzzle. The reason that it was so difficult for researchers to figure out how the brain stitches together these visual snapshots into one accurately-rendered mental representation of the visual environment is that it doesn't do that. Bruce discovered that the brain couldn't do that if it tried. The neural information that codes for saccade amplitude and direction is simply too inaccurate to determine exact relative locations of each fixation. Rather than the perception of space constancy being the result of an internal representation, Bruce determined that it is the result of a brain that simply assumes that external space remains constant, and it rarely checks to verify this assumption. In our extension of Bridgeman's formulation, we suggest that objects in the world often serve as their own representations, and cognitive operations can be performed on those objects themselves, rather than on mental representations of them.

1. Introduction

As primate eyes are known to make a series of brief fixations on different regions of the visual field, it had remained something of a mystery as to how it is that we perceive the external world as a constant volumetric space, instead of perceiving it as a series of unconnected snapshots collected by each eye fixation. A number of perceptual psychologists had suggested that a late stage of visual processing might piece together those stored snapshots from visual fixations into an internally represented mosaic of the entire visual scene (e.g., Breitmeyer, Kropfl, & Julesz, 1982; but see MacKay, 1973). According to that account, the reason we perceive space as constant is because we have an internal mental representation of the full 3D visual environment. However, Bruce Bridgeman's early training with ecological psychologist, James J. Gibson, planted the seeds for him to achieve a groundbreaking insight into how this neural mechanism of space constancy actually works: It doesn't. Throughout his career, Bridgeman and his colleagues developed psychophysical experiments demonstrating that the signals to and from the eye muscles simply aren't precise enough to allow for an accurate patching-together of the internal mosaic. Bridgeman suggested instead that visual perception may begin anew with each movement of the eyes. Steering away from the cognitivist habit of positing detailed internal representations, and instead giving a nod

* Corresponding author at: Cognitive and Information Sciences, University of California, Merced, Merced, CA 95340, United States.
E-mail address: spivey@ucmerced.edu (M.J. Spivey).

https://doi.org/10.1016/j.concog.2018.04.003
Received 31 January 2018; Received in revised form 23 March 2018; Accepted 2 April 2018

to ecological perception, Bridgeman and colleagues proposed that the brain may not need to store a detailed accurate internal model of the visual scene, and instead may rely heavily on the visual environment itself to function as the memory that subserves a perception of space constancy. Bridgeman's psychophysical observations played a pivotal role in inspiring research on change blindness (Simons & Rensink, 2005), on the role of eye movements in natural tasks (Hayhoe, Bensinger, & Ballard, 1998), and even on the spatially-extended self (Spivey & Spevack, 2017). Bridgeman's careful laboratory experimentation deserves considerable credit for the fact that numerous fields have now begun to move away from theories that place all mental activity inside the brain, and instead embrace theories of perception and cognition that rely on physical parameters of the environment as fundamental components of mental computation. But letting go of the sovereign importance of mental representations in describing how visual perception works has not been an easy task for most cognitive and perceptual scientists.

2. Complete mental representations of visual input

When one looks around the visual environment, one typically experiences a sense of *space constancy*: the perception that the world is a stable three-dimensional whole surrounding one's body – not a fragmented collection of disconnected images, or a changing mosaic of snapshots. This perception of space constancy is somewhat surprising when one examines exactly how the eyes work when they look around that environment. Firstly, the tiny central portion of the retina, the fovea, is the only place where vision has high acuity, and it covers only about one degree of visual angle. Secondly, the eyes typically make quick saccadic jumps across several degrees of visual angle every few hundred milliseconds. Therefore, it stands to reason that the visual input data that the eyes are collecting, while they look around the environment, is indeed a fragmented collection of disconnected images. How does the brain convert those fragmented snapshots into a perception of space constancy?

That question vexed perceptual psychologists – including a young Bruce Bridgeman – throughout the 1970s and 1980s. Operating on the assumption that in order to perceive something one must have an internal representation of that something, a number of researchers proposed that the disconnected snapshots were stitched together into one spatially-integrated whole (e.g., Breitmeyer et al., 1982; but see MacKay, 1973). According to this proposal, the reason we perceive our visual environment as a constant whole is because we construct an internal representation of a constant whole. This spatial integration process is sometimes compared to Marr's (1982) proposed three stages of vision. David Marr was fully aware of the imperfections in how primate eyes collect visual data, therefore he referred to the early stages of visual information processing as generating "sketches," but he suggested that the final stage was a fully fleshed out 3-D model of the scene. Marr called his first stage of visual information processing a "primal sketch" where a variety of visual features are extracted by feature detectors. Neurophysiological and psychophysical data have generally proven consistent with the existence of something like this feature-extraction stage (Morgan, 2011; Felleman & Van, 1991; but see Olshausen & Field, 2005). Marr called his second stage of visual information processing a "2½-D sketch," treating it as an intermediate stage where visual features are combined to identify surfaces, contours and rudimentary object shapes, without a fully realized spatial organization of them. Again, neurophysiological and psychophysical data have generally proven consistent with the existence of something like this surface, contour, and crude object processing stage (Lee, Mumford, Romero, & Lamme, 1998; Tanaka, 1997). Finally, Marr called his third stage of visual information processing a "3-D Model" of the scene, where viewpoint-invariant representations of the visual objects in the environment are reconstructed from the information in the 2½-D sketch. This is where his theory may deviate from the facts. While this putative 3-D Model stage could, in principle, provide some explanation for why people feel as though they visually perceive a complete and stable three-dimensional environment surrounding them, neurophysiological and psychophysical data have failed to find evidence for this third and final processing stage. It looks as though the human brain simply may not have a subsystem that generates a viewpoint-invariant 3-D Model of objects in the visual environment (see Edelman & Intrator, 2003; Tarr & Bülthoff, 1998).

However, in the 1980s, there was a real push to identify the mechanisms that build up this proposed 3-D Model in the mind. In one infamous experiment, human subjects viewed a partial grid of dots peripherally for a couple hundred milliseconds, then made an eye movement to that grid's location, where it was replaced by a new partial grid of dots for a mere 17 ms. If the 3D model of the visual environment was cognitively constructed by fusing together images collected over different eye fixations – so called trans-saccadic integration – then subjects should be able to fuse the visual memory of their first-viewed partial grid with their second-viewed partial grid, and thus accurately report which single dot was missing from their combined union. Initial reports from this experiment indicated that people were indeed able to perform this task far more accurately than a control condition that did not involve eye movements. But Bruce Bridgeman's research career at that time had already taught him that an experimental result like this was highly unlikely. Bruce had helped identify the mechanisms of saccadic suppression, whereby visual processing is largely shut down throughout the duration of a saccadic eye movement (Bridgeman, Hendry, & Stark, 1975). Thus, the reported integration of visual information from immediately before, and immediately after, a saccade sounded surprising to him. When Bruce failed to replicate the original experiment with a P4 phosphor screen (on which luminance decays in microseconds), he was able to determine that the original experiment likely suffered from an artifact of phosphor luminance from the first grid image persisting during presentation of the second grid image (Bridgeman & Mayer, 1983; see also Irwin, Yantis, & Jonides, 1983; O'Regan & Levy-Schoen, 1983; Rayner & Pollatsek, 1983). Basically, in the original flawed experiment, the "visual memory" was happening on the screen itself, not in the participants' brains.

But if a 3D Model of the scene is not being constructed in the mind, how is it that we feel as though we perceive our visual environment as one constant whole? How can we possibly perceive something for which there is no internal mental representation? As it turns out, the theoretical framework of ecological perception has, for decades, paved the way in explaining how perception can take place *without* detailed internal representations. In fact, while an undergraduate student at Cornell University, Bruce Bridgeman