Vision Research 116 (2015) 36–44

Contents lists available at ScienceDirect

Vision Research

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

# Transformation priming helps to disambiguate sudden changes of sensory inputs

Alexander Pastukhov[a,b,c,*], Solveiga Vivian-Griffiths[b,c,d], Jochen Braun[b,c]

[a] Department of General Psychology and Methodology, Otto-Friedrich-Universität, Bamberg, Germany
[b] Center for Behavioral Brain Sciences, Magdeburg, Germany
[c] Cognitive Biology, Otto-von-Guericke Universität, Magdeburg, Germany
[d] School Psychology, Cardiff University, UK

ARTICLE INFO

*Article history:*
Received 15 April 2015
Received in revised form 6 September 2015
Accepted 19 September 2015
Available online 29 September 2015

*Keywords:*
Transformations
Structure-from-motion
Shape-from-shading
Streaming-bouncing
Priming
Attention
Visual perception
Perceptual inference
Visual memory
Prior knowledge

ABSTRACT

Retinal input is riddled with abrupt transients due to self-motion, changes in illumination, object-motion, etc. Our visual system must correctly interpret each of these changes to keep visual perception consistent and sensitive. This poses an enormous challenge, as many transients are highly ambiguous in that they are consistent with many alternative physical transformations. Here we investigated inter-trial effects in three situations with sudden and ambiguous transients, each presenting two alternative appearances (rotation-reversing structure-from-motion, polarity-reversing shape-from-shading, and streaming-bouncing object collisions). In every situation, we observed priming of transformations as the outcome perceived in earlier trials tended to repeat in subsequent trials and this repetition was contingent on perceptual experience. The observed priming was specific to transformations and did not originate in priming of perceptual states preceding a transient. Moreover, transformation priming was independent of attention and specific to low level stimulus attributes. In summary, we show how "transformation priors" and experience-driven updating of such priors helps to disambiguate sudden changes of sensory inputs. We discuss how dynamic transformation priors can be instantiated as "transition energies" in an "energy landscape" model of the visual perception.

## 1. Introduction

Our brain must reconstruct the outside visual world from a sensory evidence that is always incomplete and is always intrinsically ambiguous (Gregory, 2009; Metzger, 2009; Yuille & Kersten, 2006). To make things worse retinal input constantly changes due to self-motion, changes in illumination, object-motion, etc. This poses an enormous challenge, as very different physical changes can produce identical changes in sensory evidence. An object changing its size (an inflated frog) and an object getting closer (you are walking toward the frog) could produce the same change in sensory evidence (a change in the size of a retinal image) (Combe & Wexler, 2010; Koenderink, 1986). A change of a retinal projection's shape may imply that object moved (a leaf was moved by a wind), that you moved (you walked past the tree), or some combination of both (a leaf was moved by a wind as you were walking past the tree) (Wexler, Panerai, Lamouret, & Droulez, 2001). An activation-pattern of cone cells on the retina that corresponds to somebody's face may change because the person blushed (surface has changed) or because the person stepped from a direct sunlight into an ambient illumination in the shadow or because a cloud obstructed the sun (illumination has changed) (Jameson & Hurvich, 1989). Ambiguity of change in sensory evidence makes it hard for the perceptual system to identify a unique physical cause and to determine whether constancy of a particular visual feature must be maintained. Yet, this unique physical cause is all that matters and is what our visual system is trying to correctly represent in perception.

None of the examples above correspond to a rare and exceptional event. On the contrary, they are the norm for the dynamic environment that we actively explore and which is full of objects, animals, clouds, wind, etc. It is the ubiquity of these events that raises the question of how the visual system resolves their dynamic ambiguity. The general answer to the problem is to gather and exploit prior knowledge (Friston, Breakspear, & Deco, 2012; Gregory, 2009; Metzger, 2009; Yuille & Kersten, 2006), and this process has been well studied from both behavioral (Kristjánsson

---

* Corresponding author at: Department of General Psychology and Methodology, Otto-Friedrich-Universität Bamberg, Markusplatz 3, 96047 Bamberg, Germany.
*E-mail address:* Pastukhov.Alexander@gmail.com (A. Pastukhov).
*URL:* http://www.uni-bamberg.de/allgpsych (A. Pastukhov).

http://dx.doi.org/10.1016/j.visres.2015.09.005
0042-6989/© 2015 Elsevier Ltd. All rights reserved.

& Campana, 2010; Pastukhov & Braun, 2011, 2013b) and theoretical (Friston et al., 2012; Pastukhov, García-Rodríguez, et al., 2013) perspectives, even though the neural implementations are still poorly understood (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). The main focus of prior research was the knowledge about physical states (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Weiss, Simoncelli, & Adelson, 2002; Yang & Purves, 2003; Yuille & Kersten, 2006), however this type of knowledge serves only as a weak constraint because the number of transformations by far outstrips the number of states.

Accordingly, our visual system also relies on the knowledge about physical transformations (in addition to, and independent of, the similar information on physical states) to determine the most likely cause of a change in sensory evidence. Because of that in examples above certain transformations are more likely to be perceived than other. Previous work of on transformation priors (Barbur & Spang, 2008; Combe & Wexler, 2010; Pastukhov, Vonau, & Braun, 2012; Tse, 2006; Tse & Logothetis, 2002; Wexler & van Boxtel, 2005) demonstrated their importance to the dynamic perception and their link to ecological constraints of the outside world. Present work extends this by asking the question whether this prior knowledge is gathered from the recent perceptual experience or can be considered to be static.

## 2. Materials and methods

### 2.1. Observers

All participants had normal or corrected-to-normal vision. Observers were naive to the purpose of the experiments and were paid for their participation. Procedures were approved by the medical ethics board of the Otto-von-Guericke Universität, Magdeburg "Ethik-Komission der Otto-von-Guericke-Universität an der Medizinischen Fakultät" and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

### 2.2. Apparatus

Stimuli were generated with MATLAB using the Psychophysics Toolbox (Brainard, 1997). Stimuli were displayed on a CRT screen (Iiyama VisionMaster Pro 514, iiyama.com) with a spatial resolution of 1600 × 1200 pixels and refresh rate of 100 Hz. The viewing distance was 73 cm, so that each pixel subtended approximately 0.019°. In all experiments, background luminance was kept at 36 cd/m$^2$. The experimental room was lit dimly (ambient luminance at 80 cd/m$^2$).

### 2.3. Stimuli and procedure

#### 2.3.1. Experiment 1a. Structure-from-motion

Eighteen observers participated in the experiment. Structure-from-motion (SFM) stimulus (see Fig. 1A and Movies 2–3) consisted of 100 dots distributed randomly over the surface of the illusory sphere. The diameter of the sphere was 5°, rotation period 0.2 Hz. The diameter of a single dot was 0.057°, luminance – 110 cd/m$^2$. The dots were semi-transparent, i.e. the luminance of the overlap was a sum of individual luminance levels. This provided no cue on which dot is "on top" during the overlap to exclude any possible occlusion effects. Individual trials consisted of a random stimulus onset delay (0.5–1 s, drawn from a uniform distribution), presentation interval (1.5 s) and response interval (unspeeded response, mean duration 587.9 + 19.4 ms), see Movies 2–3. Planar motion of all dots was inversed at a random time-point $T_{\text{change}}$ between 0.5 s and 1 s after the stimulus onset (drawn from a uniform distribution), see Fig. 1D and Movies 2–3. Observers used arrow keys to report on the initial and the final direction of illusory rotation. Observers reported unclear/mixed percept by pressing the "down" arrow key (2.31 ± 0.63% of total trials). Each block contained 40 On- and Off-intervals (400 trials per observer).

#### 2.3.2. Experiment 1b. Shape-from-shading

Nine observers participated in the experiment. Shape-from-shading (SFS) stimulus (Fig. 1B and Movie 4) had outer diameter of 2° and inner diameter of 0.7°, gradient rings had width of 0.3°. Stimulus orientation was defined in the direction of gradient. The display in Fig. 1B corresponds to the orientation of 90°. Individual trials consisted of a random stimulus onset delay (0.5–1 s, drawn from a uniform distribution), presentation interval (1.5 s) and response interval (unspeeded response, mean duration 969.9 ± 164.6 ms). The initial orientation of the display was pseudo-randomly selected from a uniform distribution with a 22.5° step. The display was rotated by 180° at a random time-point $T_{\text{change}}$ between 0.5 s and 1 s after the stimulus onset (drawn from a uniform distribution), see Fig. 1E and Movie 4. Observers reported on the initial and final state of the perceived shape using arrow keys (*up* – concave, *down* – convex). Observers reported unclear/mixed percept using a "left" arrow key (3.39 ± 1.34% of total trials). A single experimental session consisted of twelve blocks. Each block contained 64 On- and Off-intervals (768 trials per observer, 64 trials per orientation).

#### 2.3.3. Experiment 1c. Streaming-bouncing

Nine observers participated in the experiment. Streaming-bouncing (SB) stimulus (Fig. 1C and Movie 5) consisted of two symmetric trapezoid objects with identical height and bottom sides (both 2°) but different upper sides (0.8° and 1.2°). Objects moved with a speed of 7°/s along linear trajectories, so that they crossed behind a circular occluder (Ø3°, total presentation duration 1.5 s), see Fig. 1F and Movie 5. Response was unspeeded, mean duration 351.9 ± 62.7 ms. In half of the trials (selected randomly), objects continued the linear motion, whereas in the other half of the trials they "bounced" off each other. Observers used arrow keys to report whether they perceived streaming (objects continued the linear motion, *left arrow*) or bouncing (objects "bounced" off each other, altering their motion path, *right arrow*). A single experimental session consisted of ten blocks. Each block contained 80 On- and Off-intervals (800 trials per observer).

#### 2.3.4. Experiment 2

Nine observers participated in the experiment. Procedure was identical to that of Experiment 1a. Display in the baseline condition was identical to that in Experiment 1a. In the second condition, the planar motion inversion was omitted on half of randomly selected trials, producing a nearly unambiguous perception of stable illusory rotation.

#### 2.3.5. Experiment 3a. Specificity to location

Six observers participated in the experiment. Structure-from-motion (SFM) stimulus was identical to that used in the main experiment. Procedure was identical to that of Experiment 1a except for the location of the display. It was presented 2.5° to the left or to the right off the fixation. Location was altered on every trial, initial location at the beginning of the block was randomized.

#### 2.3.6. Experiment 3b. Specificity to axis of rotation

Six observers participated in the experiment. Structure-from-motion (SFM) stimulus was identical to that used in the main experiment. Procedure was identical to that of Experiment 1a except for the axis of rotation of the display. It was presented as rotating either around a vertical or around a horizontal axis. The