Vision Research 49 (2009) 2131–2139

Contents lists available at ScienceDirect

Vision Research

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

How long do intrinsic and extrinsic visual cues take to exert their effect on the perceptual upright?

Bahar Haji-Khamneh *, Laurence R. Harris

Dept. Psychology, York University, 4700 Keele St., Toronto, ON, Canada M3J 1P3

ARTICLE INFO

Article history:
Received 10 October 2008
Received in revised form 14 April 2009

Keywords:
Orientation perception
Scene perception
Object perception
Visual frame
Horizon

ABSTRACT

We determined the amount of time it took for intrinsic and extrinsic visual cues to determine the perceptual upright. The perceptual upright was measured using a probe, the identity of which depended on its perceived orientation (the Oriented Character Recognition Test). A visual background that filled the field of view and contained both intrinsic and extrinsic cues was presented in different orientations and for presentation times of between 50 and 500 ms followed by a mask. The contribution of each class of cue was identified by exploiting their different degrees of ambiguity. Intrinsic cues include scene structure (e.g., walls, floor and ceiling of an indoor scene) which indicates four potential up directions, and the horizon which indicates two possibilities. Extrinsic cues, which rely on information not in the image such as a surface acting as a support structure for an object, signal the direction of up unambiguously. The contribution of each class of visual cue could thus be identified from the number of cycles its effect showed as the background was presented in all orientations round the clock. While the more high-level extrinsic cues to up exerted a larger influence on the perceptual upright than the intrinsic cues, the magnitude of each cue's effect increased with presentation time at approximately the same rate with a time constant of about 60 ms. This finding poses a challenge for bottom–up theories of scene perception and suggests that low-level and high-level information are processed in parallel at least insofar as they indicate orientation.

1. Introduction

Vision tells us about the identity of objects ('seeing') but also carries proprioceptive information about the body's orientation relative to the world. Orientation is fundamental to perception and the recognition of objects depends on their orientation. The perceived direction of 'up' has conventionally been measured using the subjective visual vertical (e.g., Mittelstaedt, 1983). However, the orientation at which objects appear upright (the perceptual upright) is not always the same as the orientation of the subjective visual vertical because the perceptual upright is more heavily influenced by orientation of the visual background. Dyde, Jenkin, and Harris (2006) define the perceptual upright (PU) as being the orientation at which objects are recognized as being "the right way up". The right way up is the orientation at which objects are most readily and accurately identified and is fundamental to our ability to interact with the environment. The perceptual upright is conceptually distinct from the 'canonical orientation' which defines 'the right way up' as the orientation at which objects are most accurately and speedily recognized (see for e.g., Jolicoeur, 1985; McMullen & Jolicoeur, 1992). While the perceptual upright and the 'canonical orientation' are closely related concepts, and would likely both be influenced to the same extent by background scene orientation, the canonical orientation is derived from reaction time data whereas the PU is derived from a character recognition task. The perceptual upright is derived from a combination of visual and vestibular cues, together with an internal representation of the orientation of the body (Asch & Witkin, 1948a; Dyde et al., 2006; Mittelstaedt, 1986, 1999). Here we investigate specifically the contribution of the visual cue to the perceptual upright.

A typical scene contains both intrinsic and extrinsic visual cues to orientation. The overall frame or structure of the scene (floor or ground plane, walls, ceiling or sky) and the orientation of the horizon (even if not directly visible) are intrinsic to a scene. By contrast, the spatial-relationships between and within objects (that a table can act as a support surface for an object; that a lampshade is at the top of a lamp standard) are not intrinsic to scenes and have to be learned through familiarity with statistical regularities in the environment (Schwarzkopf & Kourtzi, 2008) and an internalization of the laws of physics (McIntyre, Zago, Berthoz, & Lacquaniti, 2001). These learned relationships constitute an axis of polarity that does not change when the overall scene changes in orientation. Such extrinsic cues will be referred to as polarizing cues. The knowledge that light comes from above (Mamassian & Goutcher, 2001; Ramachandran, 1988) can also be used to specify the orientation of an object or scene using shading and shadows. The interpretation of this cue can be altered by experience suggesting

* Corresponding author. Fax: +1 416 736 5814.
E-mail address: bahar.haji@gmail.com (B. Haji-Khamneh).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.06.003

that the light cue is also at least partially extrinsic (Adams, Graf, & Ernst, 2004). Whether intrinsic and extrinsic cues are processed by the same or different mechanisms is unknown.

Intrinsic and extrinsic cues both contribute to determining the PU. However, as Fig. 1 shows, some of these cues have different degrees of ambiguity and indicate more than one direction of up. The fact that different cues are differentially ambiguous can be used to identify their contributions in a given scene. The intrinsic cue that comes from the structure of a room provides four potential directions of up: as the scene is rotated, each of these directions aligns with gravity every 90° of rotation. Likewise, the line specifying the elevation of the horizon simultaneously indicates two directions of upright separated by 180°. In contrast to these ambiguous intrinsic cues, extrinsic cues specify a unique direction of up. Each of these cues is able to influence the orientation of the PU. Thus when a scene filling the visual field is presented at all orientations, the effect induced by the three classes of visual components within it can be distinguished by the number of cycles of shift of the perceptual upright that the tilted scene induces: the effect of the frame cues will complete four cycles, the horizon's effect will complete two and extrinsic cues will always indicate a unique direction.

While much is known about various properties of the global context such as color (Oliva & Schyns, 2000; Steeves et al., 2004) and spatial frequency (Rousselet, Joubert, & Fabre-Thorpe, 2005), relatively little is known about the influence of the orientation of the global context on the perception of self and object orientation (Rousselet, Macé, & Fabre-Thorpe, 2003; Vuong, Hof, Bülthoff, & Thornton, 2006). Extracting the gist of a scene can be done in less than 150 ms (Hegde, 2008) but is the time it takes to extract a gist comparable to the time it takes for a scene to exert an influence on the perception of objects within it? Here we measured the time course with which each class of cue present in the scene exerted its effect, expecting that differential processing systems would be reflected in different amounts of time needed for each type of cue to exert its effect. If higher-level extrinsic polarizing cues require more semantic and spatial processing than relatively low-level frame and horizon cues, then we might expect that such cues would exert their effect at a later stage than low-level intrinsic cues and should take longer. Conversely, if low-level and high-level information were processed in parallel, we would expect no differences in the time course of intrinsic and extrinsic cues.

To test these hypotheses we used the Oriented CHAracter Recognition Test (OCHART) (Dyde et al., 2006) which exploits the notion that the letters 'p' and 'd' rely on their orientation for their identity. By identifying the orientation at which the letter's identity is least certain (i.e., when either identify is equally likely to be perceived) we can obtain an estimate of the orientation at which its orientation is most certain: the perceptual upright. The influence of the orientation of the visual background was obtained by repeating OCHART with the background at different orientations. Each background was presented for a fixed period of time between 50 and 500 ms followed immediately by a pattern mask that limited the processing time to the presentation duration.

2. Methods

2.1. Subjects

Three females and five males between the ages of 24 and 45 participated in these experiments. All observers had normal or corrected-to-normal vision. All observers gave informed consent as required by the Ethics Guidelines of York University which complies with the 1964 Declaration of Helsinki. Six of the participants were volunteers and the other two were compensated at a rate of $10 per session. All participants took part in all experiments.

2.2. Apparatus

Stimuli were presented on a 21 in. Dell P1110 Trinitron monitor with a resolution of 28.3 pixels/cm and a mean luminance of 43.15 cd/m^2 at a refresh rate of 120 Hz (i.e., 8.33 ms/frame). Stimuli were composed one frame at a time and presented using Psyscope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993; MacWhinney, Cohen, & Provost, 1997). Because the timing of the stimulus and mask presentation on the computer screen was

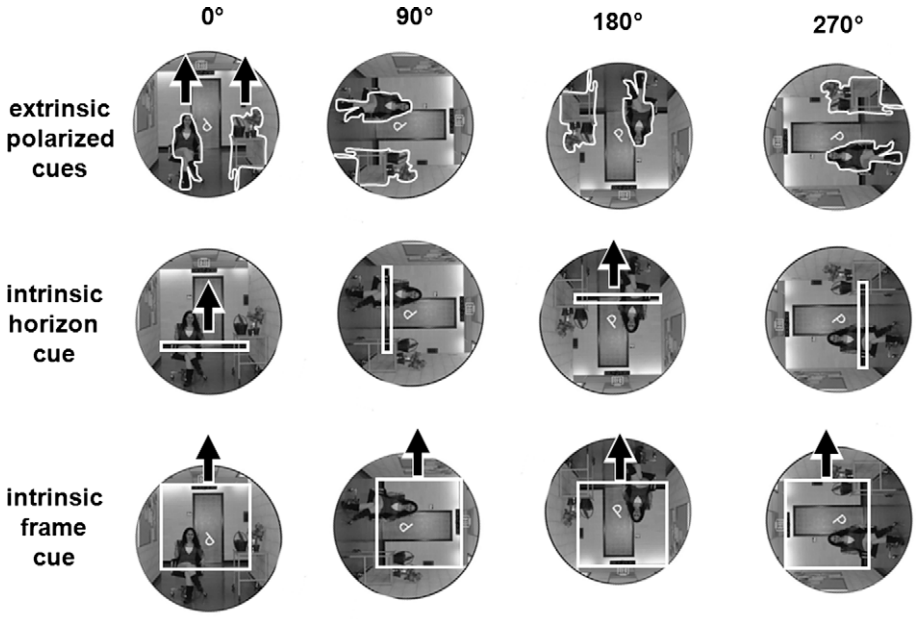

Fig. 1. A visual scene contains several cues to orientation including high-level extrinsic polarizing cues (highlighted in the top row) and low-level intrinsic cues from the horizon and visual frame (highlighted in the middle and bottom rows, respectively). When the picture is rotated through 360°, the direction of up specified by the polarizing cues rotates through one cycle (top row); the direction specified by the horizon cue rotates two cycles (middle row) and the direction indicated by the frame cue (the square formed by the edges of the walls and the floor and ceiling) rotates through four cycles (bottom row).

critical for this experiment we verified the timing of the stimuli carefully. Stimuli were presented for periods of time that were multiples of the frame duration and this was confirmed using a light-sensitive diode pointing at the screen. The screen was viewed at a distance of 25 cm through a black circular shroud that obscured peripheral vision and that reduced the viewing area to a circle subtending 28.5° of visual arc (Fig. 2).

2.3. Test for perceptual upright

The Oriented CHAracter Recognition Test (OCHART) technique exploits the fact that the perceived identities of some objects depend solely on their orientation (Dyde et al., 2006). The probe we used was the ambiguous character . The character subtended approximately 3.1° × 1.9° of visual arc. Its perceived identity as a 'p' or 'd' is based exclusively on its orientation.

The method of constant stimuli was used to find the two orientations where the character was equally likely to be perceived as a 'p' or a 'd'. The character was presented six times at 24 different orientations spanning the range from 0° to 345° in 15° increments. The bisector of the two orientations at which the character was maximally ambiguous was taken as the orientation at which its identity was maximally certain and defined as the perceptual upright.

2.4. Background stimuli

The character probe was superimposed on a 28.5° circular background picture which was a colored photograph of a scene that was rich in intrinsic and extrinsic visual cues specifying up (insert to Fig. 3). Since scenes with man-made structures include more vertical lines and hence stronger intrinsic cues than natural scenes (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007) we chose a photograph that was taken indoors. The background image was presented at 16 orientations spaced equally around the clock (i.e., in steps of 22.5°). Thus there were 408 (16 × 24) probe/background combinations which were each presented six times in a randomized order with presentation times of 50, 75, 150 and 500 ms (6, 9, 18 and 60 frames, respectively) resulting in a total of 9,216 (16 backgrounds × 24 probe orientations × 4 presentation times × 6 repetitions) presentations. These were completed in six blocks of 1536 trials each. Each block was approximately 1 h long and the participants were allowed to take breaks. No feedback was given. In order to keep the relative amount of intrinsic and extrinsic cues constant, the same background photograph was used throughout. Any possible adaptation effects on the PU were assessed by comparing data taken from the first half of the data collection period with those taken in the second half.

Fig. 2. Subjects viewed the display through a shroud that obscured all peripheral vision, masked the screen to a 28.5° diameter circle and set the viewing distance at 25 cm. Participants responded using the buttons on a game pad, visible on the desk.

2.5. Procedure

Participants were instructed to identify the probe symbol as a 'd' or a 'p' using the left and right buttons respectively on a game pad. They were seated at a desk with their heads firmly positioned against the circular tunnel approximately 25 cm from the computer screen (Fig. 2). Participants pressed any button on the keyboard to start the experiment. At the start of each trial a 0.45° fixation point appeared against a grey background and stayed on for 100 ms (12 frames) after which a probe/background stimulus combination was presented for either 6, 9, 18 and 60 frames. The probe/background stimulus was followed immediately by a structure mask for 100 ms (12 frames). The mask was followed by a grey screen at which time observers pressed the button to indicate the perceived identity of the symbol ('p' or 'd'). After the participant responded, the fixation point came on again and the next trial commenced. The sequence is summarized in Fig. 3.

2.6. Calculating the perceptual upright

The percentage of times the symbol was identified as a 'p' was plotted as a function of the probe orientation for each background orientation and presentation time. An example is shown in Fig. 4. Two cumulative Gaussian functions were fitted to the participants' responses to determine each of p–d and d–p transition orientations. The cumulative Gaussians were defined as:

$$y = \frac{100}{1 + e^{-(x-x_0)/b}}\% \quad (1)$$

where x_0 corresponds to the 50% point and b is the standard deviation. The 50% point corresponds to the orientation of the probe at which either interpretation was equally likely i.e., the 'transition orientation'. The orientation midway between the p–d and d–p transition angles was taken as the PU. In the example shown in Fig. 4 this is at 2.1°. The mean of the standard deviations of each of the two cumulative Gaussians was taken as the standard deviation of the observer's response in each testing condition.

2.7. Testing the time it took to identify the character probe

To establish that the probe could be identified at the different presentation times used in this study we measured identification performance for a range of presentation times. Participants sat in the equipment as for the experiment proper (Section 2.5) and viewed the fixation point. They were then shown the probe against the neutral grey background in either the 'p' or 'd' orientation for between 2 and 9 frames (16.7–75 ms) followed by a structure mask for 100 ms. Participants were asked to identify the letter using the left and right buttons on the game pad. The next trial commenced immediately after they responded. Ten 'p's and 10 'd's were presented at each of the eight timing conditions resulting in 160 (20 × 8) trials in total. The presentation of all stimuli was randomized. No feedback was given. This phase of the experiment was completed in approximately 5 min.

2.8. Convention

The orientations of all probe and background stimuli are defined with respect to the body mid-line of the observer. Zero degree refers to the orientation of the body axis. Positive orientations are clockwise ('rightwards') relative to this reference

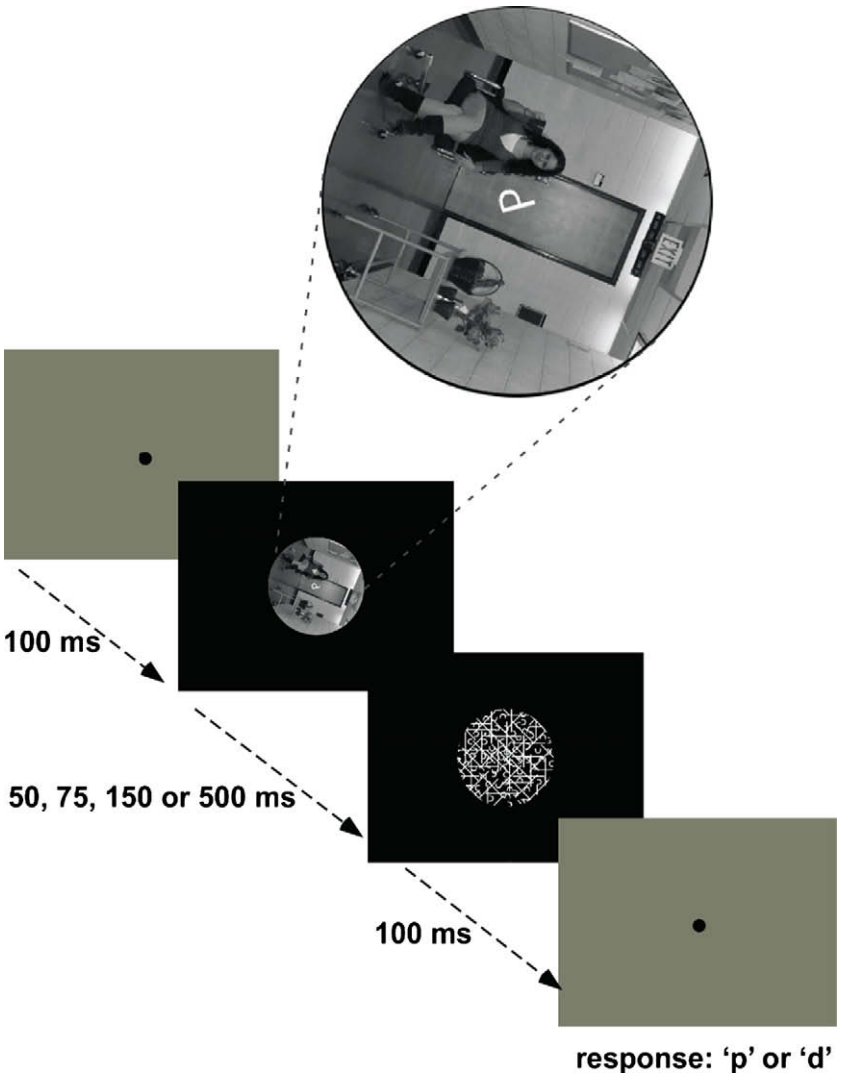

Fig. 3. The sequence of events in a typical trial. After fixation point offset, participants saw the stimulus, a composite of background and probe, for 50, 75, 150 or 500 ms. Stimulus offset was followed by a structure mask for 100 ms and then a grey screen with the fixation point. The background image was presented in color.

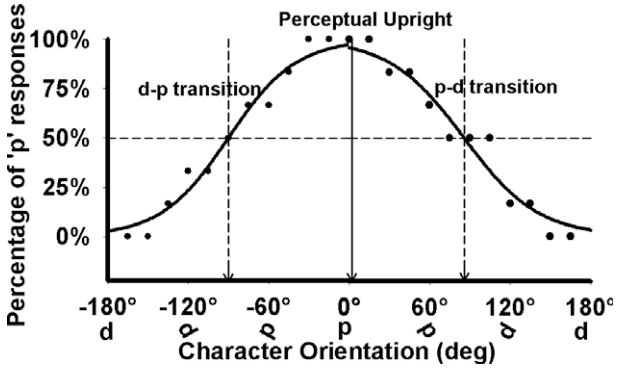

Fig. 4. Typical psychometric functions obtained from a single background orientation (in this case upright) and presentation time (in this case 150 ms). The percentage of times the character was identified as a 'p' is plotted against its orientation. Cumulative Gaussians were plotted through the data from which the two points of maximum ambiguity (the 50% point) were found (in this case, at −90.1° and +85.8° indicated on the graph by vertical dashed arrows). The perceptual upright is defined as being half way between these orientations (2.1° in this example, illustrated by the solid arrow).

orientation, negative orientations are counter-clockwise ('leftwards'), as seen by the observer. The 'p' symbol is described as being 0° when the vertical shaft of the symbol is aligned with the body axis with the letter bowl to the right (i.e., when the character was presented as an upright 'p').

3. Results

3.1. Character identification

To establish that the probe could be identified at the different presentation times used in this study we measured identification performance for a range of stimulus onset asynchronies (SOAs). The percentage of accurate character identification was calculated for each SOA for each participant. The average across all participants is plotted as a function of the stimulus–mask onset asynchrony (t) in Fig. 5. A two-parameter cumulative Gaussian function on a 50% pedestal

$$y = 50 + \frac{50}{1 + e^{-(t-t_{75})/b}}\% \quad (2)$$

was fitted to these data where t_{75} corresponds to the presentation time at which subjects were 75% correct threshold and b is the standard deviation. The mean presentation time at which participants were able to correctly identify the character 75% of the time was

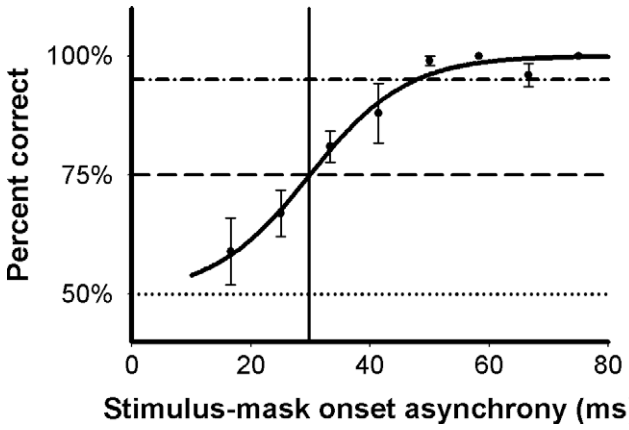

Fig. 5. Time to identify the probe character. The percentage of times the probe was correctly identified is plotted as a function of stimulus–mask onset asynchrony. 75% correct performance was reached at 29.6 ms as indicated by the vertical line. Data points represent the mean proportion of correct responses over 20 trials averaged across all eight participants. The 50%, 75% and 95% responding levels are indicated by horizontal dashed lines.

29.6 ms as shown by the vertical line in Fig. 5. Participants were capable of identifying the probe 95% of the time at the shortest OCHART viewing condition used for the main experiment (50 ms).

3.2. The effect of visual background orientation on PU

The orientation of the PU is plotted as a function of the orientation of the visual background for each of the four stimulus–mask onset asynchronies in Fig. 6. The open symbols with standard error bars are the average of the data from all eight participants. The PU was strongly influenced by the orientation of the visual background, varying by more than ±16° at all stimulus–mask onset asynchronies. The solid line fitted through the data is the model fit (see below, Section 3.4).

3.3. The effect of adaptation on the PU

The effects of adaptation on PU were tested by analyzing data from the first and second half of each participant's data session separately and comparing the just noticeable differences (jnd) and the points of maximum ambiguity of the first sample with that of the second sample. The early and late samples were also compared to the overall data with respect to these parameters. For the early and late samples, neither the just noticeable difference of the psychometric judgments nor the points of maximum ambiguity were significantly different from each other ($p = .88$ and .93, respectively). Likewise, the early and late samples did not differ significantly from the overall data either with respect to their b values ($p = .75$) or with respect to their points of maximum ambiguity ($p = .83$).

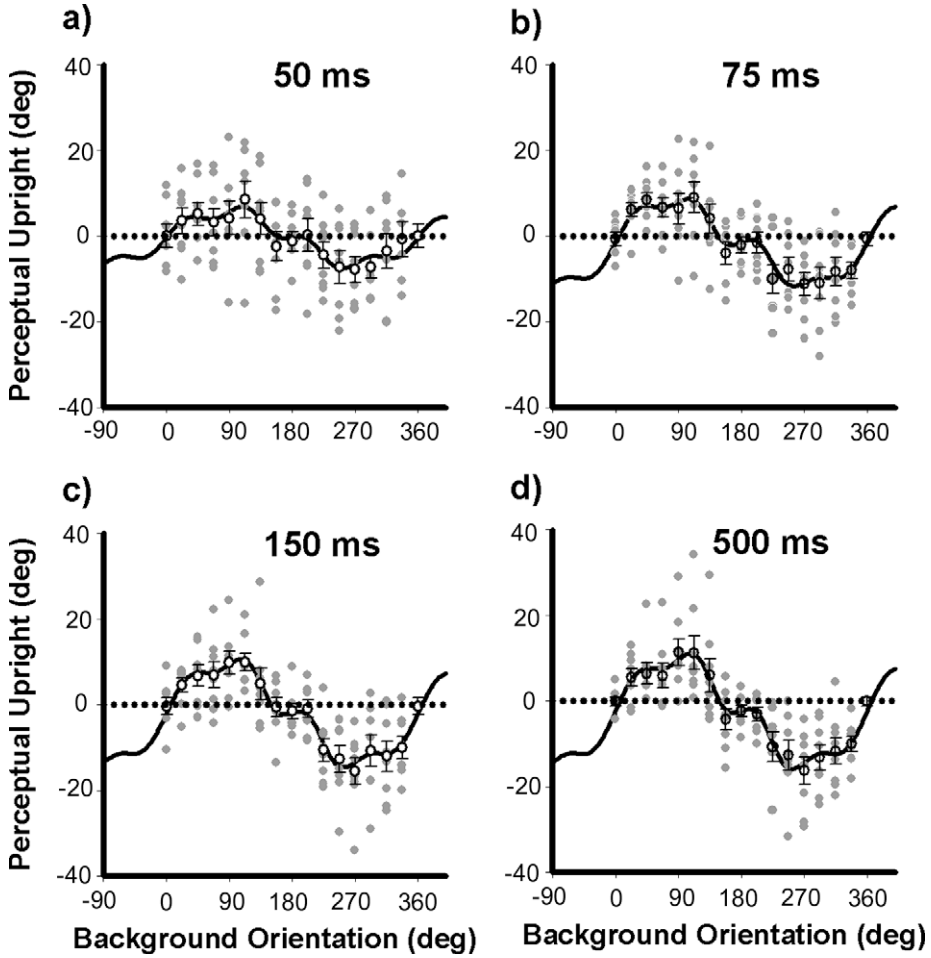

Fig. 6. This figure illustrates how the orientation of the PU (vertical axis) changes with the orientation of the background scene (horizontal axis) for stimulus–mask onset asynchronies of (a) 50 ms, (b) 75 ms, (c) 150 ms and (d) 500 ms. The grey dots represent the PU at each background orientation for each participant (obtained as shown in Fig. 4). The means are depicted by the open circles with corresponding standard errors. The solid curves are the lines of best fit of the vector sum model (see text and Fig. 7).

3.4. Vector model

We modeled the effect of visual cues on the PU using the weighted vector model described in Dyde et al. (2006). In this model the orientation of the body, gravity, and visual cues are treated as vectors in their veridical directions with lengths proportional to their relative weights. The orientation of the PU is then predicted from the vector sum of these three vectors. We extended this model by breaking down the visual vector into its individual components, namely the extrinsic polarizing cues, the horizon cues, and the frame cues. The model does not include the light-from-above direction as a separate vector. This extrinsic cue was present in our background scene (Fig. 3) and its contribution would be part of the 'polarizing cue'. The vector that represents the direction indicated by the extrinsic polarizing cues corresponds to the orientation of the visual background. Therefore, this vector aligns itself with gravity only once as the background makes one revolution. The horizon cue however, aligns twice as the background makes each revolution and the frame cue aligns four times (see Fig. 1). Since observers were tested only in the upright body orientation, the body vector and the gravity vector were always aligned. For the present study we therefore treated these two as a single vector of unity length and expressed the lengths of the other three vectors relative to this gravity/body axis vector. Each vector length was then converted into a percentage of the sum total of all the vector lengths for that condition. This model, with three free parameters corresponding to the lengths of the three visual vectors, was then fitted to the data describing the orientation of the PU for each background orientation using an established optimization algorithm (the Marquardt-Levenberg technique, see Press, Flannery, Teukolsky, & Vetterling, 1988). Three parameters were thus obtained for each participant and each SOA. The solid line plotted through the data in Fig. 6 shows the output of this model plotted through all the data for each presentation time. The regression coefficients were 0.14, 0.36, 0.56 and 0.68 for the 50, 75, 150 and 500 ms SOAs, respectively.

The mean amplitude of each of the three visual vectors obtained for each subject and each SOA was compared using repeated-measures ANOVA with the visual cues (polarizing vs. horizon vs. frame) as the within-subjects factors. The main effect of visual component on vector length was significant $F(1.073, 3.219) = 51.793$, $p = .004$. Paired sample t-tests revealed that while the mean vector length for polarizing cues was significantly different from the mean vector length for the horizon cues $t(3) = 8.217$, $p = .004$ as well as the frame cues $t(3) = 6.693$, $p = .007$, the weightings of the latter cues were not significantly different from each other $t(3) = -.37$, $p = .731$.

3.5. Presentation time and the influence of visual cues on PU

In order to describe the amount of time that each component needed to exert its effect on the perceptual upright we fitted an exponential growth function to the length of each visual vector for each participant plotted as a function of stimulus–mask onset asynchrony (Fig. 7a). Time constants were obtained for exponential growth of each visual component and averaged across participants as summarized in Fig. 7b. A repeated-measures ANOVA comparing the time constants of extrinsic polarizing cues ($M = 68.7$ ms, 95% confidence interval = 33–103 ms), horizon cues ($M = 51.7$, 95% confidence interval = 16–86 ms) and frame cues ($M = 58.7$, 95% confidence interval = 26–90 ms) revealed no significant difference ($F(1.36, 9.56) = .062$, $p = .878$), suggesting that all functions increased at approximately the same rate with an overall mean time constant of 59.7 ms.

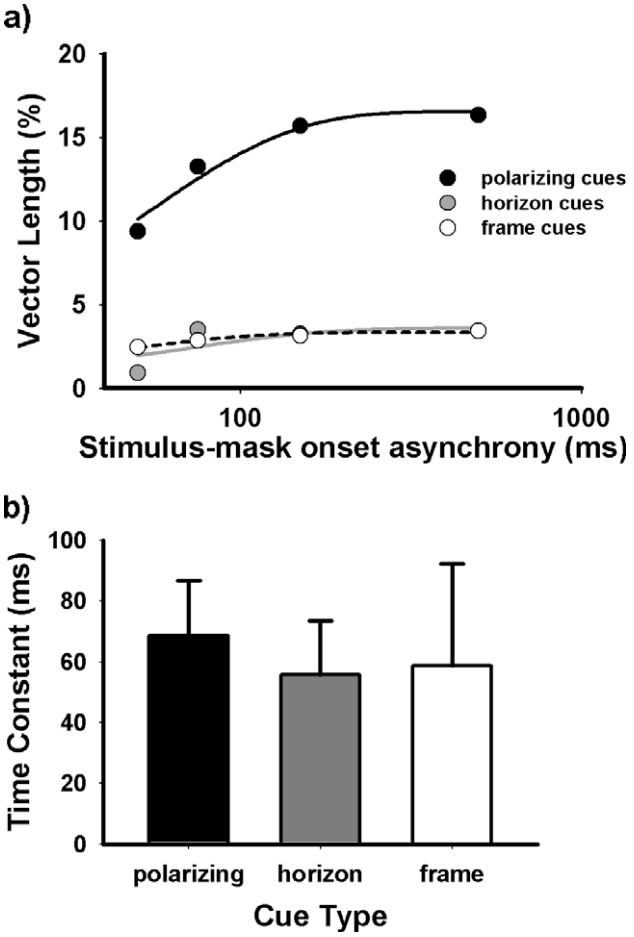

Fig. 7. (a) The length of each of the three visual vectors are plotted as a function of stimulus–mask onset asynchrony for each stimulus–mask onset asynchrony. They are modeled with exponential fits shown as a black line through filled data points for polarizing cues, grey line through grey data points for horizon cues and dashed line through open data points for frame cues. The standard errors for these data points (from 0.016 to 0.02) were too small to be depicted graphically. (b) The mean time constants were averaged across participants ($n = 8$) and shown here as a bar chart with standard error bars. The mean time constants were 68.9, 51.7 and 58.7 ms for polarizing, horizon and frame cues, respectively. The 95% confidence intervals for the time constants were 33–103 ms for the polarizing cues, 16–86 ms for the horizon cues and 26–90 ms for the frame cues. No significant difference was found between time constants.

4. Discussion

4.1. Summary

The background image evoked shifts in the perceptual upright (PU) linked to each of the three classes of visual orientation cues contained in the picture: extrinsic polarizing cues and intrinsic horizon and frame cues. The polarizing cues had the most influence on the orientation of the PU as evidenced by the longer length of the vector associated with this cue. The different extent to which each visual cue affects the PU is likely to at least partially depend on the content of the visual image. If there are fewer structural cues visible, for example, then it is likely they will have a relatively smaller effect. However, the relative lengths of the vectors do not concern us here. What is interesting is that the time it took for each component to exert its effect was the same for all three cues with a time constant around 60 ms (Fig. 7b). This is much longer than the time it took to reliably identify the letter probe (about 30 ms). The fact that all three components took about the same amount of time to exert their effect was surprising to us. We had expected the

intrinsic polarizing cues to take longer, based on the presumed higher-level of processing involved.

The jnd and the points of maximum ambiguity were not affected as a result of adaptation to the single background scene that was used in this experiment. Thus the influence of each cue on the PU did not change as a result of adaptation. However, the use of only a single image raises other concerns with regards to the generalizability of our results especially that of the absolute values of our obtained time constants. Familiarity with visual stimuli has been shown to change processing (Müller, Metha, Krauskopf, & Lennie, 1999) and it is a plausible conjecture that processing speed may differ in more natural visual environments (where every image is analyzed anew). It is important to keep this in mind for the remainder of the discussion.

4.2. The processing time of visual orientation information

The time that the visual stimulus was available was limited by the use of a pattern mask. Visual processing is a dynamically changing phenomenon (VanRullen & Thorpe, 2001) and pattern masking has been an invaluable tool in examining the various levels of information processing (Breitmeyer & Ögman, 2006) for over 100 years (Exner, 1868). Pattern masking refers to a masking method in which the target stimulus spatially overlaps with and precedes the mask and the visibility of the target is limited by its temporal proximity to the mask (Kahenman, 1968). This method has been used to study the recognition of simple geometric forms and faces (Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005), to assess the time it takes to categorize (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005) or recognize (Rieger, Braun, Bülthoff, & Gegenfurtner, 2005) a scene. Here we applied pattern masking to the use of orientation information. We adopted the notion that psychophysical applications of masking have traditionally assumed: that the mask essentially erases and/or adds noise to the target stimulus at early pre-cortical visual areas to effectively terminate its further processing (e.g., Carrasco, Williams, & Yeshurun, 2002; Rauschenberger & Yantis, 2001; Rieger et al., 2005). That is, we make the assumption that there exists an early representation of the stimulus in a sensory buffer that encodes the physical attributes of the stimulus (Marr, 1982). This template is constantly processed by higher cognitive centers at increasingly abstract levels (Lamme & Roelfsema, 2000; Rieger et al., 2005). The mask would disrupt processing by replacing the template of the target with its own. The stimulus–mask onset asynchrony is the critical variable. We conclude from our experiments that the processing time (taken as the time constant of growth) of each visual component is approximately the same at around 60 ms: that is that the scene needs to be visible for about 60 ms (a stimulus–mask onset asynchrony of 60 ms) for both intrinsic and extrinsic cues to be available. The fact that there was no systematic difference between them suggests parallel extraction of the three components, since a serial extraction (in which higher-level cues are constructed from lower-level ones) would have resulted in cumulative time effects with higher-level cues taking the time it took to process the lower-level ones plus the time it took to combine those cues.

4.3. Comparison with other timings in vision

Fast and accurate object recognition is critical to our survival in the real world. An object is typically embedded in a global scene containing many other objects as well as other features. Much research has focused on elucidating the time course of object recognition (Biederman, 1972; Mumford, 1994; Rao & Ballard, 1999; Rosch, 1976) as well as that of scene perception (Biederman, Mezzanotte, & Rabinowitz, 1982; Hegde, 2008; Joubert et al., 2007). Taken together, these studies suggest that within as little as the first 150 ms from stimulus onset, the visual system has already extracted enough meaningful information from complex scenes to be able to perform highly demanding tasks such as categorization (Rieger et al., 2005; Rousselet et al., 2003; Thorpe, Fize, & Marlot, 1996). Our findings suggest that the effects of the scene on object orientation may be even speedier: within only 60 ms from stimulus onset enough information has been extracted from the visual scene to influence the perception of object orientation.

What kind of information can be adequately extracted in such a short time interval as to have an impact on our visual orientation? Scene processing is extraordinarily fast. Twenty-four milliseconds of undistorted processing provides sufficient information to recognize scenes above chance level and after 90 ms perfect recognition accuracy is reached (Rieger et al., 2005). Ringach, Hawken, and Shapley (1997) examined the temporal dynamics of the orientation tuning of V1 cells in anesthetized monkeys. They found that while orientation tuning was first observed 30–45 ms after stimulus onset, it improved over an additional 40–85 ms during which time tuning became progressively sharper. This range is compatible with our intrinsic time constants (of 51.7 and 58.7 ms). While this coincidence may explain the magnitude of the time constants for the frame and the horizon components, polarized cues include not only complex objects but also complicated and learned relationship between objects. The orientation tuning of V1 cells cannot be the sole underlying mechanism for their effects.

Our time constants are also in accordance with results of studies of higher-level object categorization. For example, 40–60 ms of stimulus duration is sufficient to allow for accurate categorization of scenes (Bacon-Macé et al., 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Loschky et al., 2007). Fei-Fei et al. (2007) used a free-recall task to probe the nature of information that is accessible during scene categorization. They found that, while the viewers' open-ended descriptions of briefly flashed scenes were dominated by low-level sensory descriptors at short stimulus onset asynchronies, subjects used more object-related and semantic language to describe what they saw at longer presentation times. The transition occurred somewhere between 40 and 67 ms. Other researchers found a similar temporal window for maximal scene categorization accuracy between 40 and 81 ms using an ERP paradigm (Bacon-Macé et al., 2005). Rieger, Koechy, Schalk, Grueschow, and Heinze (2008) found that at least 50 ms of undistorted information accumulation is necessary for the effect of scene-object orientation incongruence on reaction time to emerge. They deemed it likely that this effect emerges at the junction in processing when information regarding both orientation (De Caro & Reeves, 2000, 2002; Hamm & McMullen, 1998) and scene semantics (Ganis & Kutas, 2003; Henderson & Hollingworth, 2003) first became accessible. Taken together, these studies suggest that our time constants of 50–70 ms may reflect a surprisingly speedy object categorization mechanism which carries sufficient information, including that pertaining to orientation and semantics, to identify an object's polar axis and specify the extrinsic cues to orientation. However, whether this finding holds for natural processing of non-familiar scenes is an empirical question.

4.4. The importance of the background for object processing

A visual scene contains intrinsic and extrinsic cues that provide orientation information and also define what kind of scene it is. The relative contribution of higher-level extrinsic and lower-level intrinsic orientation cues in determining a scene, and its subsequent effect on the perception of objects within the scene, is not established. Low-level properties of scenes such as color (Delorme, Richard, & Fabre-Thorpe, 2000; Gegenfurtner & Rieger, 2000) and orientation (Rousselet et al., 2003; Vuong et al., 2006) do not seem capable of disrupting object classification. Rousselet et al. (2003)

found only marginal effects of scene inversion on face/animal classification for briefly flashed scenes and Vuong et al. (2006) found no such effects on the detection of humans in a scene. By contrast, a recent study Rieger et al. (2008) found that scene rotation (but not inversion), as well as incongruence between object and scene orientation, had an inhibitory effect on object classification. Further they found that the processing of objects embedded within a scene was disrupted when the object was upright and the scene was rotated, suggesting that object processing relies heavily on scene orientation. In the same vein, Dyde et al. (2006), using the OCHART letter probe, and Mittelstaedt (1999) and (Asch & Witkin, 1948a, 1948b) using a rod, showed that scene orientation can affect the perceived orientation of embedded objects. These effects of the visual background represent composite effects of all the components, both high-level (extrinsic) and low-level (intrinsic) of the visual scene. The present study represents the first attempt to isolate the individual contributing components within a natural scene.

Traditionally, scene recognition has been thought of as the culmination of a bottom–up process of information extraction. According to this proposition, early low-level modules such as contour detection, shade perception, and stereo perception, are integrated to give rise to individual objects. These objects are in turn combined to give rise to high-level scene recognition (Bülthoff & Mallot, 1987; Driver & Baylis, 1996; Hildreth & Ullman, 1993; Marr, 1982; Nakayama, He, & Shimojo, 1995). Sensory- or feature-level properties of a scene, such as shape, are perceived faster than spatial-relationships and semantic-level information (Fei-Fei et al., 2007). The present finding corroborate a parallel-processing model of scene processing since lower-level information such as the structural frame of the scene and the horizon were found to be processed at the same speed as the higher-level polarized objects.

4.5. Conclusions

In summary, we have shown that all three visual cues to orientation are processed at approximately the same rate (50–70 ms) at least for a familiar image. This processing time range may reflect an exceedingly rapid categorization mechanism in the case of the polarized cues, and low-level orientation tuning in the case of the frame and the horizon cues. That all cues are processed at the same speed contradicts theories that claim object perception precedes scene perception and instead our data support parallel-processing of objects and contexts. Moreover, despite their equivalent processing speed, polarized cues are the most salient cue in determining the perceptual upright. That all of this information is extracted from the visual environment and perhaps assessed for usefulness in less than 100 ms, implies that the brain has the extraordinary capability of dealing with a vast amount of visual information and clutter in just a glance. Further experiments are required to illuminate the speed and nature of processing of extrinsic and intrinsic cues in more natural visual environments. Also, the relative contribution and speed of processing of the light-above-prior warrants further investigation.

Acknowledgments

These experiments were supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Space Agency (CSA). We would like to give special thanks to Richard Dyde and Michael Barnett-Cowan for their help in setting up and running this project.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057–1058.

Asch, S. E., & Witkin, H. A. (1948a). Studies in space orientation. 1. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology, 38*(3), 325–337.

Asch, S. E., & Witkin, H. A. (1948b). Studies in space orientation. 2. Perception of the upright with displaced visual fields and with body tilted. *Journal of Experimental Psychology, 38*(4), 455–477.

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorization. *Vision Research, 45*(11), 1459–1469.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception – Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177.

Breitmeyer, B., & Öğman, H. (2006). *Visual masking: Time slices through conscious and unconscious vision* (2nd ed.). New York: Oxford University Press.

Bülthoff, H. H., & Mallot, H. A. (1987). Interaction of different modules in depth perception. In *First international conference on computer vision*, London, England (pp. 295-305).

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision, 2*, 1–42.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope – An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods Instruments and Computers, 25*(2), 257–271.

De Caro, S. A., & Reeves, A. (2000). Rotating objects to determine orientation, not identity: Evidence from a backward-masking/dual-task procedure. *Perception and Psychophysics, 62*(7), 1356–1366.

DeCaro, S. A., & Reeves, A. (2002). The use of word–picture verification to study entry-level object recognition: Further support for view-invariant mechanisms. *Memory and Cognition, 30*(5), 811–821.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research, 40*(16), 2187–2200.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology, 31*(3), 248–306.

Dyde, R. T., Jenkin, M. R., & Harris, L. R. (2006). The subjective visual vertical and the perceptual upright. *Experimental Brain Research, 173*(4), 612–622.

Exner, S. (1868). Über die zu einer gesichtswahrnehmung nöthige zeit [On the time necessary for visual perception]. *Wiener Sitzungsber Math-Naturwiss Cl Kaiserlichen Akad Wiss, 58*(2), 601–632.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123–144.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10*(13), 805–808.

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology – Human Perception and Performance, 24*(2), 413–426.

Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84*(4), 405–439.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 356–377). New York, NY, US: Oxford University Press.

Hildreth, E. C., & Ullman, S. (1993). The computational study of vision. In E. C. Hildreth & S. Ullman (Eds.), *Foundations of cognitive neuroscience* (pp. 581–630). Cambridge, MA: MIT Press.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition, 13*(4), 289–303.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*(26), 3286–3297.

Kahenman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70*(6, Pt. 1), 404–425.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feed-forward and recurrent processing. *Trends in Neuroscience, 23*, 571–579.

Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research, 45*(17), 2287–2297.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology – Human Perception and Performance, 33*(6), 1431–1450.

MacWhinney, B., Cohen, J., & Provost, J. (1997). The PsyScope experiment-building system. *Spatial Vision, 11*, 99–101.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81*(1), B1–B9.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman and Co.

McIntyre, J., Zago, M., Berthoz, A., & Lacquaniti, F. (2001). Does the brain model Newton's laws? *Nature Neuroscience, 4*(7), 693–694.

McMullen, P. A., & Jolicoeur, P. (1992). Reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology – Human Perception and Performance, 18*(3), 807–820.

Mittelstaedt, H. (1983). A new solution to the problem of the subjective vertical. *Naturwissenschaften, 70*(6), 272–281.
Mittelstaedt, H. (1986). The subjective vertical as a function of visual and extraretinal cues. *Acta Psychologica, 63*(1–3), 63–85.
Mittelstaedt, H. (1999). The role of the otoliths in perception of the vertical and in path integration. *Annals of the New York Academy of Sciences, 871*, 334–344.
Müller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (1999). Visual adaptation in visual cortex to the structure of images. *Science, 285*, 1405–1408.
Mumford, D. (1994). Large-scale neuronal theories of the brain. In C. Koch & J. L. Davis (Eds.), *Neuronal architectures for pattern-theoretic problems* (pp. 125–152). Cambridge, MA, US: The MIT Press.
Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual cognition: An invitation to cognitive science*, (2nd ed., Vol. 2, pp. 1–70). Cambridge, MA, US: The MIT Press.
Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C: The art of scientific computing*. New York: Cambridge University Press.
Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(6152), 163–166.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87.
Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature, 410*(6826), 369–372.
Rieger, J. W., Koechy, N., Schalk, F., Grueschow, M., & Heinze, H. (2008). Speed limits: Orientation and semantic context interactions constrain natural scene discrimination dynamics. *Journal of Experimental Psychology – Human Perception and Performance, 34*(1), 56–76.
Rieger, J. W., Braun, C., Bülthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *Journal of Vision, 5*(3), 275–286.
Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*(6630), 281–284.
Rosch, E. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.
Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition, 12*(6), 852–877.
Rousselet, G. A., Macé, M. J. M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*(6), 440–455.
Schwarzkopf, D. S., & Kourtzi, Z. (2008). Experience shapes the utility of natural statistics for perceptual contour integration. *Current Biology, 18*(15), 1162–1167.
Steeves, J. K. E., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *Journal of Cognitive Neuroscience, 16*(6), 955–965.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.
VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13*(4), 454–461.
Vuong, Q. C., Hof, A. F., Bülthoff, H. H., & Thornton, I. M. (2006). An advantage for detecting dynamic targets in natural scenes. *Journal of Vision, 6*(1), 87–96.