Are face representations depth cue invariant?

The visual system can process three-dimensional depth cues defining surfaces of objects, but it is unclear whether such information contributes to complex object recognition, including face recognition. The processing of different depth cues involves both dorsal and ventral visual pathways. We investigated whether facial surfaces defined by individual depth cues resulted in meaningful face representations—representations that maintain the relationship between the population of faces as defined in a multidimensional face space. We measured face identity aftereffects for facial surfaces defined by individual depth cues (Experiments 1 and 2) and tested whether the aftereffect transfers across depth cues (Experiments 3 and 4). Facial surfaces and their morphs to the average face were defined purely by one of shading, texture, motion, or binocular disparity. We obtained identification thresholds for matched (matched identity between adapting and test stimuli), non-matched (non-matched identity between adapting and test stimuli), and no-adaptation (showing only the test stimuli) conditions for each cue and across different depth cues. We found robust face identity aftereffect in both experiments. Our results suggest that depth cues do contribute to forming meaningful face representations that are depth cue invariant. Depth cue invariance would require integration of information across different areas and different pathways for object recognition, and this in turn has important implications for cortical models of visual object recognition.

Introduction

Despite continuous change to the retinal image of objects around us, the transformation of the retinal image into a complex object representation remains constant and tolerant to changes in viewpoint, size, or retinal location (Kobatake & Tanaka, 1994; Li, Cox, Zoccolan, & DiCarlo, 2009; Tanaka, 1996; Vuilleumier, Henson, Driver, & Dolan, 2002). This transformation is thought to arise from hierarchical processing of edge information in the two-dimensional (2-D) retinal image by the ventral visual pathway and is effectively captured by computational models such as HMAX (X. Jiang et al., 2006; Konen & Kastner, 2008; Riesenhuber & Poggio, 1999, 2000). Object representations derived in this manner are also tolerant to viewpoint, size, and position changes without implicitly or explicitly defining a three-dimensional (3-D) object representation. This is surprising considering that we can recognize complex objects with depth cues such as texture and structure from motion (SFM; Farivar, Blanke, & Chaudhuri, 2009; C. H. Liu, Collin, Farivar, & Chaudhuri, 2005), suggesting that independent of size, position, and viewpoint tolerance, visual object representations ought to also be depth cue invariant (Farivar, 2009).

While we can recognize objects from pure depth cues, it is unclear to what extent depth information can drive object recognition (i.e., category level or individualization as well). A related problem is that individual depth cues are processed by different mechanisms: Binocular disparity and SFM are understood to be extracted by mechanisms in the dorsal visual pathway, while shading and texture appear to heavily engage areas in the ventral visual pathway (Cant, Large,
McCall, & Goodale, 2008; Farivar, 2009; Georgieva, Todd, Peeters, & Orban, 2008; Grunewald, Bradley, & Andersen, 2002; Janssen, Vogels, & Orban, 2000; Y. Liu, Vogels, & Orban, 2004; Nelissen et al., 2009; Vanduffel et al., 2002). Depth information must be integrated into a single representation in the late stages of the ventral visual pathway in order to allow for object recognition (Farivar, 2009; Farivar, Blanke, et al., 2009; Janssen et al., 2000; Nelissen et al., 2009), but such an integrative process would violate models of ventral visual hierarchy. Disparity and SFM appear to be extracted only at the late stages of the dorsal pathway, and their relay to the late ventral pathway areas would violate the strict ventral stream hierarchy assumed by current object recognition models (Farivar, 2009).

Here we sought to address the first problem outlined above—the extent to which depth information can contribute to object recognition at the individuation level. In our previous studies, we found that subjects could recognize unfamiliar faces from individual depth cues such as texture or SFM in a 1:8 matching task (Farivar, Blanke, et al., 2009; C. H. Liu et al., 2005). While these results could be taken as evidence for depth-driven face individuation, it is also possible that subjects did not utilize a face representation per se to perform the task but rather used local feature matching or shape analysis. In the four studies described here, we utilized the face identity aftereffect to more directly assess the potential of pure depth cues for forming face representations and the efficacy of these 3-D driven face representations in bringing about adaptation effects within and across depth cues.

Face identity can be represented in relation to a prototypical face and in a multidimensional face space (Blanz, O'Toole, Vetter, & Wild, 2000; Leopold, O'Toole, Vetter, & Blanz, 2001). If we consider a single face being defined as a point in a multidimensional face space with the “average” face resting at the origin, then one can extrapolate a face that is opposite of a given face. Extended viewing of such an “antiface” results in the perception of the target identity even when looking at an average face. This face identity aftereffect has been used effectively to test the invariant properties of face recognition and the mechanisms underlying face coding in adults and children (Jeffery & Rhodes, 2011; F. Jiang, Blanz, & O'Toole, 2007; Leopold et al., 2001). Instead of using synthetic or human-made objects of which the dimensions are unknown (and their representation is therefore difficult to assess), we used the face identity aftereffect to probe complex object representation from 3-D depth cues.

Most previous studies on face recognition relied mainly on 2-D images, and only a few studies examined the contribution of 3-D information to face recognition. Prior work using face identity aftereffect suggests that face representations are invariant to changes in viewpoint and illumination (F. Jiang et al., 2007; F. Jiang, Blanz, & O’Toole, 2009), supporting the notion that a 3-D representation is formed to achieve this tolerance. Addition of binocular disparity information results in higher accuracy of face recognition, which also aids in generalization across different viewpoints (Chelnokova & Laeng, 2011; C. H. Liu, Chai, Shan, Honma, & Osada, 2009; Schwaninger & Yang, 2011), suggesting that 3-D cues can aid object recognition. Whether faces defined by pure depth cues can create complex object representations such as that of a face remains an open question. We therefore measured the face identity aftereffect with 3-D stimuli constructed from individual depth cues to address this issue.

We reasoned that if faces defined by single depth cues result in a meaningful face representation—that is, a face representation that maintains the relationship of facial identities in a multidimensional face space—then adaptation to antifaces defined by the same depth cue ought to result in face identity aftereffects. Furthermore, if these face representations are depth cue invariant, then the antiface adaptor and the test face do not need to be defined by the same depth cue for the effect to hold. Using precisely controlled 3-D stimuli devoid of 2-D contaminants, we describe four experiments that together show (a) that face representations can be formed from pure 3-D depth cues and (b) that these representations are depth cue invariant.

### Experiments 1 and 2: Face Identity Aftereffect From Individual Depth Cues

The purpose of Experiments 1 and 2 was to determine whether the perceptual representation of faces from individual depth cues is comparable with that of face photographs. Face images can be effectively identified even when individuating information is restricted by morphing the face to an average face. By creating a gradation of identity information in faces, we can measure a face identification threshold—that is, the degree of facial information needed to accurately identify a face. Face identification thresholds are shifted by adaptation to an antiface, or a face with features and qualities opposite of a target identity. We used this face identity aftereffect to probe whether 3-D facial surfaces tap into a similar multidimensional representation as that shown for face photographs (F. Jiang et al., 2007, 2009; Leopold et al., 2001).

In Experiment 1 we determined the face identification thresholds based on participants’ responses to faces at individual morph levels using the method of...
constant stimuli. Face identification thresholds were measured across four participants, each responding in 3,360 trials carried out over 23 hr. To better assess the generalizability of the adaption effect, in Experiment 2 we shortened the study duration by implementing an adaptive staircase method and further constraining the design. We tested a further set of 18 participants in this constrained version of the procedure. To assess the depth cue invariance of face representations, we then carried out two additional studies utilizing method of constant stimuli (Experiment 3) and adaptive staircase (Experiment 4) with adaptor and test stimuli rendered with different depth cues.

**Experiment 1**

**Method**

**Participants**

Four healthy subjects participated in this part of the study. All had normal or corrected-to-normal vision. The mean age of the participants was 25 years (ages 23–26 years; one female and three males). Before the start of each experiment, subjects became familiar with the experimental paradigm and different depth cues defining the facial surfaces by being provided with examples for each depth cue. This study was approved by the Montreal Neurological Institute Ethics Board.

**Stimuli**

The stimuli used in these studies consisted of 3-D facial surfaces varying in identity information compared with an average face and defined by a single depth cue devoid of any 2-D or 3-D contaminants. **Facial surfaces**: Synthetic 3-D facial surfaces were generated using FaceGen Modeller 3.5 (Singular Inversions, Inc., Toronto, Canada). Four individual facial surfaces (two males, two females; Figure 1A) were generated randomly using the “generate” feature of this software. Using a custom application provided by the vendor, antiface adaptors were generated by mirroring the multidimensional representation of the target face at the general, gender-neutral average face. The resulting antiface has features that are contrary to the target face while being uninformative of the target face (Blanz et al., 2000). We then generated identity morphs from −30% to 100% identity in steps of 10% (i.e., 30% away from the mean toward the antiface to 100% away from the mean toward the target face; Figure 1B).

Facial surfaces were edited in a professional 3-D modeling package (3d Studio Max 2013, Autodesk, San Rafael, CA) to minimize the utility of the face size and external contour in the identification tasks. All faces and morphs were first resized in height to match the dimensions of the average facial surface. To obstruct the utility of the facial contour in a way that generalized across all our viewing conditions, we added depth noises to the outer edges of the facial surfaces, which effectively generated random contours for each stimulus.

**Rendering of facial surfaces by pure depth cues**: Surfaces were defined purely by one of four different 3-D depth cues: shading, texture, SFM, and binocular disparity. We used a professional rendering package (3d Studio Max 2013, Autodesk, Inc.) to render the shaded, texture, and disparity-defined stimuli and a custom-made algorithm to render the SFM stimuli (Figure 2).

Shaded stimuli were rendered in 3d Studio Max. A directional light source was introduced at a 45° angle with respect to the horizontal at the top of the face in frontal view in the scene. All other cues (glossiness, specular level, and softness level) were set to zero. The frontal view of the face was rendered in orthographic projection.

The texture stimuli were generated in a manner identical to that used in C. H. Liu et al. (2005) using a procedural fractal noise texture map in 3d Studio Max 2013. Glossiness, softness level, and specular level were set to zero. A procedural fractal noise texture was used with the following parameters: high threshold = 0.75, low threshold = 0.25, noise level = 3, and noise size = 20 units. The face stimuli were also set to the maximum self-illumination (100%), which eliminated shading, causing the 3-D depth to be defined purely by the texture gradient.

Motion-defined facial surfaces were generated in the same manner as in Farivar, Nelissen, and Vanduffel (2009) by projecting 2-D random dot patterns to the facial surface, rotating the 3-D surface in depth by 0.5° in each frame, and displacing the dots to maintain a uniform dot density (Farivar, Nelissen, et al., 2009). Overall, there were 34 frames for each facial surface in which the 3-D face rotated in depth from −4° (toward left) to 4° (toward right) from a central plane. A total of 12,000 white dots were projected onto the surface, and the antialiased dot diameter was 3 pixels against a black background.

Disparity-defined surfaces were generated using a high-density regular noise texture rendered with two cameras simulating the position of the two eyes at different viewing angles (−2.86° and +2.86°). The high density of the texture ensured that the stimuli could be seen only binocularly. The positions of the cameras were adjusted based on the position of the eyes (60 cm) when viewing the face in full screen on the monitor. The face stimuli were set to the maximum self-illumination (100%), which eliminated shading and shadows. A high-density regular noise texture was used
with the following parameters: noise size = 1.0 unit, noise level = 3.0 units, high threshold = 0.6, and low threshold = 0.4.

The shaded, texture, and SFM stimuli were presented on a ViewSonic Professional Series monitor and an LG Electronics F900P 19-in. cathode ray tube monitor with a spatial resolution of 1280 × 1024 and a refresh rate of 85 Hz. Disparity-defined stimuli were presented on a Samsung 3-D monitor (model S23A700D) with shutter glasses operating at 120 Hz, with a spatial resolution of 1920 × 1080 and a refresh rate of 120 Hz (i.e., 60 Hz for each eye). The height of the face images on both monitors was approximately 10.5° of visual angle.

Procedure

The experiment started with a training session in which participants had to associate the identity of an individual face with a particular keyboard button. Feedback was provided during the training trials. Subjects were required to achieve 100% correct performance on 100% morphed faces and more than 90% correct performance on 60% morphed faces—only these two morph levels were provided during training. The subject’s reaction time was also recorded. The participant was allowed to proceed to the experimental trials if their maximum reaction time was less than 1 s, with a mean reaction time being below 0.5 s, in addition to meeting the percentage correct performance criteria. In order to eliminate the possibility of an afterimage, no fixation point was used during the training or experimental trials and subjects were encouraged to freely inspect the face stimuli.
The procedure for the adaptation sessions is depicted in Figure 3. In each trial of the adaptation sessions, the antiface adaptor was presented for a brief period of time followed by a blank gray screen for 17 ms. The adaptation time was 5000 ms for shaded stimuli, 15,000 ms for texture, 19,652 ms for the SFM, and 15,000 ms for stereo. The different times were selected based on pilot testing with each depth cue, which was aimed at minimizing the length of the adaptation sessions while maintaining comparable performance across individual depth cues.

Following adaptation, a test morph of random identity strength (from −30% to 100%) was shown as the test stimulus using the method of constant stimuli. The test face was shown for 200 ms for the shaded condition, 500 ms for both the texture and stereo conditions, and 3468 ms for the SFM condition. A four-alternative forced-choice task was used in which the participants were asked to determine the identity of the test stimulus. The depth cue used to define the adaptor and the test stimuli were the same in a single session.

In the matched condition, the antiface adaptor was always matched with the subsequent test face. In the non-matched condition, the identity of the antiface adaptor was not associated with the identity of the morphed face shown afterward. We also measured the psychometric functions for the face identification from 3-D depth cues in a separate session without any adaptors (no-adaptation condition).

In total, performance at each morph level was measured in 20 trials (5 per identity × 4 identities), resulting in 280 trials per adaptation condition and 840 trials per depth cue condition (280 trials × 3 adaptation conditions) for a total of 3,360 trials per subject (except for one subject who performed 2,688 trials, missing a session for the stereo disparity). The whole experiment lasted approximately 23 hr per subject, including the training period, which was spread out across multiple weeks.
Analysis

Psychometric Weibull functions were fitted to the responses with a nonparametric bootstrapping technique implemented in the Palamade Toolbox (Kingdom & Prins, 2010). The identification threshold was defined as the amount of identity information needed to achieve 72.2% performance, with identity information defined as the percentage morph between a target face and the average face. The guessing rate was fixed to 0.25, with a lapse rate of 0.001. We obtained group psychometric functions by fitting the Weibull to the aggregate data from all subjects. In order to determine whether antiface adaptation had a significant effect on the face identification threshold, a model comparison analysis was applied on the combined data set across all subjects (Kingdom & Prins, 2010). Using this analysis, the probability of the transformed likelihood ratio (pTLR) was determined for each subject and depth cue.

In addition, we made individual subject fits to measure normalized thresholds, which were defined as the ratio of the identification threshold in the matched or non-matched condition compared with the identification threshold of the unadapted condition. The normalized thresholds were compared using a single-sample t test.

Results

Baseline face identification threshold (no-adaptation condition)

As expected, participants performed best when facial surfaces were defined by shading with the mean threshold at 23% morph level. The mean identification thresholds were at 51%, 49%, and 44% morph level for texture-, motion-, and disparity-defined stimuli, respectively, suggesting comparable performance across these conditions. Informally, all participants but one performed better with disparity-defined faces than with faces defined by texture or SFM.

Adaptation to antiface (matched and non-matched conditions)

Robust face identity aftereffect was observed for facial surfaces defined by each depth cue (pTLR < 0.05) in the group average data. The aftereffect was strongest for facial surfaces defined by shading, and the results from SFM showed weaker aftereffects compared with the texture and stereo disparity cues that were quite comparable across subjects (Figures 4 and 5). On average, all stimuli but the SFM resulted in robust aftereffects, as revealed by the single-sample t test of the normalized threshold data (p < 0.05).

Thresholds for all subjects were lower in the matched condition than in the no-adaptation condition for all but the SFM condition. All subjects exhibited an approximately 10% decrease in their thresholds for all conditions except for SFM, where two out of four subjects had negligible aftereffects. Furthermore, all subjects showed higher identification thresholds for the non-matched condition compared with the matched condition.

Experiment 2

While the results from Experiment 1 using the method of constant stimuli are strongly indicative of face adaptation effects being general across different depth cues, we sought to better assess the generalizability of this finding in a larger group of subjects. Because it is infeasible to test a large number of subjects with the full design of Experiment 1, we reduced the testing time by constraining our design and using an adaptive staircase to accelerate threshold measurement. The feedback from Experiment 1 also allowed us to improve some aspects of the stimuli.

Method

Participants

A total of 18 subjects with normal or corrected-to-normal vision participated in the studies; 13 subjects completed all conditions, four completed all but the texture condition, and one completed only the texture condition. The mean age of the participants was 25.89 years (ages 20–48 years; 11 females and seven males). Before the start of the experiment, the subjects were familiarized with the experimental paradigm. Informed written consent was obtained from all subjects prior to the start of the study.

Stimuli

Three facial identities (John, Lili, and Susan) from the previous experiment were used for this part of the experiment. The same set of shaded and texture stimuli was used in this part of the study. However, we improved the vividness of motion- and disparity-defined stimuli by changing the parameters explained below in the algorithm that was used to create motion-defined stimuli in the previous part to make the new disparity- and motion-defined stimuli (Figure 6).

Motion-defined stimuli: Motion-defined facial surfaces were generated in the same manner as in Farivar, Nelissen, et al. (2009) by projecting 2-D random dot patterns to the facial surface, rotating the 3-D surface in depth by 1° in each frame, and displacing the dots to maintain a uniform dot density (Farivar,
Nelissen, et al., 2009). Overall, there were 35 frames for each facial surface in which the 3-D face rotated in depth from $-8^\circ$ (toward left) to $8^\circ$ (toward right) from a central plane. A total of 12,000 white dots were projected onto the surface, and the antialiased dot diameter was 3 pixels against a black background. In order to ensure uniform 2-D dot density on each frame, the algorithm shuffled dots from areas of high dot density to areas of low dot density. From feedback received in Experiment 1, the reshuffling of the dots was perceived as disruptive, and we sought to minimize the effect of this reshuffling by rendering four independent movies for each stimulus condition and then overlaying them by blending. This resulted in a smoother visual stimulus that was perceived as less disruptive by the participants.

Disparity-defined stimuli: Disparity-defined surfaces were generated using the same algorithm as previously described for motion-defined stimuli. A 2-D uniform-density white dot pattern was projected to the face surface. The face surfaces were then rotated in depth by $-4.5^\circ$ and $4.5^\circ$, and psychometric functions were fitted to the cumulative data across all subjects. For all different depth cues, the psychometric function of the matched condition is shifted toward the left compared with the no-adaptation condition, and the psychometric function of the non-matched condition is shifted toward the right compared with the no-adaptation condition.

Figure 4. The psychometric function of the group data for each depth cue. Panels A (shaded), B (texture), C (SFM), and D (stereo) represent the proportion of correct performance as a function of identity morph level in the matched (red), no-adaptation (black), and non-matched (blue) conditions. Each subject performed 20 trials per condition per morph level. Psychometric functions were fitted to the cumulative data across all subjects. For all different depth cues, the psychometric function of the matched condition is shifted toward the left compared with the no-adaptation condition, and the psychometric function of the non-matched condition is shifted toward the right compared with the no-adaptation condition.
4.5° angles, and the dots were displaced to maintain a uniform dot density. A total of 30,000 dots were projected onto the surface, and the antialiased dot diameter was 3 pixels. Four independent renderings of each frame (each

Figure 5. The mean normalized identification thresholds for different depth cues. Each individual's identification threshold for different conditions was normalized through dividing by the no-adaptation threshold. Error bars depict the standard error of the mean. The degree of improvement in the performance with respect to the matched condition or the worsening of the performance in the non-matched condition is compared with the baseline (no-adaptation/no-adaptation) shown as a straight line marked at 1. There was a significant reduction in the normalized identification threshold of the matched condition compared with the baseline for all depth cues except for SFM. The significance of the difference in the ratio between each condition versus the no-adaptation condition is marked by the asterisk (*).

Figure 6. Examples of different types of stimuli used in Experiments 2 and 4. Each facial surface was defined by a single individual depth cue. Here, an example of an average face defined by shading, texture, and stereo disparity is shown. (In the case of the stereo disparity condition, the two frames were presented dichoptically with a height of 18° of visual angle and at a distance of 50 cm from the monitor using polarized glasses.)

Figure 7. Schematic description of the adaptation condition for Experiments 2 and 4. This representation shows an example of an adaptation condition in which an either matched or non-matched test stimulus was shown following the adaptor in a trial. The no-adaptation condition was similar to that in Experiment 1.

with a different 2-D uniform random dot pattern) were overlaid on top of one another, with each layer set to a transparency of 33%, resulting in smooth percept.

Disparity-defined stimuli were presented on an HP 3-D monitor (model 2311gt) with a spatial resolution of 1920 × 1080 and a refresh rate of 60 Hz. Polarized glasses were used to visualize the disparity-defined stimuli. The height of the facial images on the monitor was approximately 18° of visual angle. The participant’s distance from the monitor was 50 cm.
**Procedure**

The experiment started with a training session in which participants had to associate the identity of two individual faces—chosen randomly from a set of three individual faces—with a particular keyboard button. The training session was the same as in Experiment 1 except that greater than 95% correct performance was required on 60% morphed faces before proceeding to the experimental sessions. Similar to Experiment 1, no fixation point was used during the training or experimental trials and subjects were encouraged to freely inspect the face stimuli.

The procedure for the adaptation sessions is depicted in Figure 7. The same antiface adaptor was shown for the whole duration of the adaptation session. For instance, if Lili and John were the two individual faces that the subject was trained on, anti-Lili or anti-John would have been the antiface adaptor for the entire adaptation session. During the adaptation condition, the antiface adaptor was presented for a brief period of time, followed by a blank gray screen for 17 ms and then by the test stimulus. The matched and non-matched conditions were shown randomly and interleaved in a single run. In the matched condition, the identity of the antiface adaptor was matched with the subsequent test face. In the non-matched condition, the identity of the antiface adaptor was not associated with the identity of the morphed face shown afterward. For instance, if the subject was trained on John and Lili, the antiface adaptor could have been either anti-Lili or anti-John. If anti-Lili was chosen as the adaptor, the testing stimulus could have been John (accounting for the non-matched condition) or Lili (accounting for the matched condition). We also measured the identity thresholds of the two tested individual face stimuli for each subject without adaptation.

The adaptation time was 5000 ms for shaded stimuli, 10,000 ms for texture and stereo, and 17,505 ms for the SFM. The different times were selected based on pilot testing with each depth cue, which was aimed at minimizing the length of the adaptation sessions while maintaining comparable performance across individual depth cues.

We used an adaptive staircase method (QUEST) as implemented in Psychtoolbox (Watson & Pelli, 1983). To enhance the robustness of the threshold estimation using the adaptive staircase, we obtained a preliminary estimate of thresholds after 40 trials (20 matched trials and 20 non-matched trials being randomized) and used this estimate to constrain the second staircase. The two staircases followed immediately after one another without awareness of the participant. A two-alternative forced-choice task was used in which the participants were asked to determine the identity of the test stimulus. The depth cue used to define the adaptor and the test stimuli were the same in a single session. The test face was shown for 200 ms for the shaded and texture conditions, 1000 ms for the stereo condition, and 1171 ms for the SFM condition based on pilot testing.

In total, there were 80 trials for the adaptation condition (40 matched, 40 non-matched) and 80 trials for the no-adaptation condition (40 trials for each individual face stimuli). The adaptation session for each depth cue took approximately 20 to 30 min, and the no-adaptation session for each depth cue lasted about 3 min. The whole experiment lasted approximately 2 hr.

**Results**

**Baseline face identification threshold (no-adaptation condition)**

As expected, participants performed best when facial surfaces were defined by shading, with the mean threshold at 12% morph level. The mean identification thresholds were at 42%, 29%, and 32% morph level for texture-, motion-, and disparity-defined stimuli, respectively (Figure 8).

**Adaptation to antiface (matched and non-matched conditions)**

Repeated measures analysis of variance was conducted to determine the effect of adaptation on identification thresholds in the different depth cue conditions. There was a significant effect of adaptation on the identification thresholds, $F(1, 12) = 35.89, p <$
A significant effect of depth cue on identification thresholds, $F(3, 36) = 7.66, p < 0.001$; and no significant interaction between depth cues and adaptation conditions, $F(3, 36) = 1.17, p = 0.34$.

Significant face identity aftereffect was observed for facial surfaces defined by each depth cue as revealed by paired-sample $t$ tests between matched and non-matched conditions in the group average data: shading, $t(16) = -6.87, p < 0.001$; texture, $t(13) = -2.21, p = 0.023$; SFM, $t(16) = -3.65, p = 0.0011$; stereo, $t(16) = -3.17, p = 0.003$. The aftereffect was strongest for facial surfaces defined by shading, and the results from texture showed weaker aftereffects compared with the SFM and stereo disparity cues that were quite comparable across subjects (Figure 8).

**Experiments 3 and 4: Transfer of Face Identity Aftereffects Across Depth Cues**

The results of Experiments 1 and 2 suggest that faces defined by individual depth cues can induce a face identity aftereffect. This supports the idea that cortical face representations can be formed by depth cues, but these results do not tell us whether such representations are invariant to depth cues. The aim of Experiments 3 and 4 was to investigate whether adaptation can transfer across different depth cues—depth cue invariant representations would exhibit such a transfer. To simplify the design, we used only shaded antiface adaptors due to their strong adaptation effect observed in Experiments 1 and 2 and tested the adaptation effect of a shaded antiface adaptor on stereo disparity– and SFM-defined test faces.

**Experiment 3**

**Method and analysis**

**Participants**

Three participants (one female and two males; ages 23–26 years) from Experiment 1 continued to this part of the study.
Stimuli

We used the same set of disparity- and motion-defined stimuli as in Experiment 1 to measure face identification thresholds.

Procedure

All procedures were identical to those in Experiment 1 except in the adaptation sessions, where the antiface adaptors were always defined by shading while the test stimuli were defined by either stereo disparity (shaded → stereo condition) or motion (shaded → SFM condition). Shaded antiface adaptors were chosen because the strongest aftereffect was seen for this type of depth cue.

Performance at each morph level was measured in 20 trials (5 per identity × 4 identities), resulting in 280 trials per adaptation condition and 840 trials per combined depth cue condition (280 trials × 3 conditions: no adaptation, matched, and non-matched) for a total of 1,680 trials per subject. The whole experiment lasted approximately 6 hr, including the training sessions, per subject. The analysis was identical to that in Experiment 1.

Results

We found robust transfer of face adaptation effects. Adaptation to shaded stimuli reduced face identification thresholds even when the test faces were defined by disparity or SFM (pTLR < 0.001; Figures 9 and 10). The effect was comparable between the shaded → stereo condition and the shaded → SFM condition. The mean normalized identification thresholds were significantly reduced to 0.63 and 0.62 for the shaded → stereo and shaded → SFM conditions, respectively (p < 0.02). At the individual level, all three subjects performed significantly better during the matched condition compared with the no-adaptation condition for both the shaded → stereo and shaded → SFM conditions (p < 0.04).

In the non-matched condition, each participant performed worse or comparable with the no-adaptation condition. The mean normalized identification thresholds in the non-matched condition were 1.25 and 1.14 morph levels for the shaded → stereo and shaded → SFM conditions, respectively (Figure 10).

Experiment 4

Method

Participants

Seventeen participants (11 females and six males; ages 20–48) from Experiment 2 continued to this part of the study.

Stimuli

We used the same set of disparity- and motion-defined stimuli as in Experiment 2 to measure face identification thresholds.

Procedure

All procedures were identical to those in Experiment 2 except in the adaptation sessions, where the antiface adaptors were always defined by shading while the test stimuli were defined by either stereo disparity (shaded → stereo condition) or motion (shaded → SFM condition). Shaded antiface adaptors were chosen because the strongest aftereffect was seen for this type of depth cue.

The adaptive staircase (QUEST) algorithm was also used in this part of the study to rapidly measure the identification threshold. There were two adaptation sessions, each lasting approximately 25 to 30 min. The whole experiment lasted approximately 1 hr. The analysis was identical to that in Experiment 2.

Results

Repeated measures analysis of variance revealed a significant effect of adaptation on the identification thresholds, F(1, 16) = 7.81, p = 0.013. There was no significant effect of depth cue on identification thresholds, F(1, 16) = 0.26, p = 0.62, and there was no
interaction between depth cue and adaptation conditions, $F(1, 16) = 0.18$, $p = 0.67$. Using a planned post hoc paired-sample t test, we found significant face adaptation effects. Adaptation to shaded stimuli reduced face identification thresholds even when the test faces were defined by disparity or SFM: shaded $\rightarrow$ stereo, $t(16) = -2.07$, $p = 0.028$; shaded $\rightarrow$ SFM, $t(16) = -1.97$, $p = 0.033$. The effect was comparable between the shaded $\rightarrow$ stereo and shaded $\rightarrow$ SFM conditions (Figure 11).

### Discussion

In this study, we examined the extent to which pure depth information can form complex object representations at the individuation level using face identity aftereffects. First, we found that an individual depth cue defining the facial surface is effective in driving face identification performance alone based on the results from the no-adaptation condition. This suggests that each individual depth cue contains significant information about facial identity to empower recognition. Second, adaptation effects were observed within and across depth cues, suggesting that a single face representation is achieved by individual depth cues and that this representation is depth cue invariant. Third, this is the first study to show that we can identify faces purely defined by stereo disparity, suggesting an important role for this depth cue in 3-D face representation.

These findings are concordant with prior psychophysical studies suggesting that 3-D complex object representation can indeed be formed by individual depth cues (Farivar, Blanke, et al., 2009; C. H. Liu et al., 2005). We found that (a) face representation from individual depth cues tap into similar representations as that of face photographs in that they also can exhibit face identity aftereffects, (b) representations formed from one depth cue transfer to other depth cues, and (c) each individual depth cue including binocular disparity can empower robust face recognition.

Both within and across different depth cue conditions, adaptation resulted in more than a 20% decrease in mean identification thresholds in the matched condition compared with the non-matched condition. Leopold et al. (2001) used face photographs in which the shift in the mean identification threshold was approximately 12.5% for the matched condition compared with the baseline condition. Although more adaptation effect is shown in our study compared with Leopold et al.’s (2001) study, the threshold level between Leopold et al.’s (2001) study and our study was different. Additionally, X. Jiang et al. (2006) showed that adaptation to frontal views of the face photographs improved the face identification for morph levels less than 25%, which is consistent with our study. Thus, our findings with depth-defined faces are consistent with previous studies measuring face identification with “normal” face photographic morphs.

### Depth cues in face perception and recognition

Prior psychophysical and neuroimaging studies mostly examined the role of combined depth cues or adding depth (mainly stereo disparity) in human object recognition (Ban, Preston, Meeson, & Welchman, 2012; Knill & Saunders, 2003; C. H. Liu, Collin, & Chaudhuri, 2000; C. H. Liu & Ward, 2006). For example, C. H. Liu et al. (2000) found that adding stereo disparity provided a slight advantage in face identification performance when recognizing faces with minimally informative shading. In a later study, C. H. Liu and Ward (2006) obtained similar results regarding the improvement in face recognition performance during stereo disparity compared with a monocular condition while the level of perspective transformation between the learning face and the testing face changed. Thus, 3-D information improves face recognition performance if other cues are corrupted or inaccessible. Our results suggest that, individually, depth cues can be highly informative of face identity, but we did not assess combinations of depth cues or inclusion of edge information. However, our results suggest that face representations are depth cue invariant and, combined with previous results, support the notion that depth cues contribute to face recognition when other information is lacking.

### Dorsal–ventral integration for recognition of objects from multiple depth cues

It is thought that the dorsal visual pathway, starting from the primary visual cortex and extending to the posterior parietal areas, is involved in the extraction of SFM and binocular disparity cues, while the ventral visual pathway is known to be engaged in the analysis of shading and texture cues that are important for object recognition (Cant et al., 2008; Farivar, 2009; Georgieva et al., 2008; Grunewald et al., 2002; Janssen et al., 2000; Y. Liu et al., 2004; Nelissen et al., 2009; Vanduffel et al., 2002). As a result, in order to generate a single 3-D object representation from different depth cues, integration across the two streams is postulated to be necessary (Farivar et al., 2009; Knill & Saunders, 2003; O’Toole, Roark, & Abdi, 2002). This contrasts with the strict serial hierarchical processing of objects in the ventral visual pathway (Farivar, 2009; Farivar et al., 2009; Grill-
Spector, Kourtzi, & Kanwisher, 2001; Kanwisher, 2010; Konen & Kastner, 2008; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; Kriegeskorte et al., 2003; O'Toole et al., 2002). Such an integrative process would violate current computational models of object recognition, such as HMAX, which are based on the strict hierarchical processing of objects in the ventral visual pathway (Riesenhuber & Poggio, 2000).

Although our results suggest the existence of an integrative process between the dorsal and ventral visual pathways, neuroimaging techniques provide more direct assessment of this problem. We hypothesize that the lateral occipital complex (LOC) plays an important role in integrating different depth cues processed by different neural mechanisms to form a single 3-D object representation. Prior neuroimaging studies have shown that by the level of the LOC, object representations are invariant to size, position, and viewpoint (Grill-Spector et al., 1999, 2001; Malach et al., 1995; Sawamura, Georgieva, Vogels, Vanduffel, & Orban, 2005). This suggests that the LOC is part of the mechanism that is involved in forming the 3-D object representations and may therefore contribute to depth cue invariance as is tolerant to other 2-D and 3-D transformations.

### Depth cue invariance in object recognition

Our study cannot compare the relative contribution of different depth cues to face recognition because we did not control the strength of the perceived depth from each cue. The stronger face identity aftereffect observed for the shaded cue, for example, might be due to the amount of depth information provided by this depth cue compared with the other depth cues, or it could be due to the manner in which the depth cues were rendered in our study. While it is tempting to interpret the stronger effect of shading as being particular in driving face identity aftereffects, our study cannot directly support or refute this idea. We did observe stronger aftereffects for the shaded-defined faces, and we used shaded adaptors in the cross-cue studies for that same reason: because the shaded faces seemed to result in stronger aftereffects than the other cues. It is possible that shaded faces are better at driving ventral pathway representations—putatively where face recognition would be carried out—hence their greater ability to form these representations and thus cause greater adaptation effects. Alternatively, the shaded faces may simply have given rise to stronger surface perception: Surface extracted from disparity, SFM, and texture all rely on extraction of gradients of features, while the shaded face does not. In this view, the stronger effect of shaded faces and other stimuli is not due to a common pathway between shape from shading and face identification but rather is due to a stronger surface representation for shaded stimuli. Future studies may consider the equivalency of depth information from different depth cues for a direct comparison.

Our results suggest that shading, texture, SFM, and disparity could all result in face identity adaptation, and in at least two pairings of shading–SFM and shading–disparity the adaptation transferred across depth cue. These results suggest that face representations are invariant to depth cue, and as such the results would predict that other pairings would result in aftereffects as well.

Depth cue invariance (Farivar, 2009) represents another important aspect of object recognition alongside position, size, and viewpoint invariance. A key aspect of depth cue invariance is, in our view, that it implies object representations that are 3-D in nature—surface representations that can be created by any depth cue or even implied by 2-D edge and contour information. While our results strongly suggest that face representations are depth cue invariant, we speculate that depth cue invariance is a general property of object recognition mechanisms. It remains to be seen whether depth cue invariance is present in single-cell responses in the ventral pathway, at the population level, or whether different objects by depth cue representations exist across cortical patches.

A key challenge in our studies was that the measurements of the adaptation effect were time consuming, forbidding us from testing large numbers of subjects on all conditions. Even with our reduced design utilizing the adaptive staircase, testing on single depth cues demanded 2 hr of testing, while the cross-depth cue task required an additional 1 hr. Despite these limitations, our study provides robust psycho-physical evidence for the role of individual depth cues in face identification and the depth cue invariance of face representations as measured by the face identity aftereffect.

### Conclusions

We examined whether faces defined by single depth cues result in a meaningful face representation—that is, a face representation that maintains the relationship of facial identities in a multidimensional face space. Based on our findings, 3-D complex object representations such as faces can be driven by an individual depth cue defining the object’s surface at the individuation level. We then asked whether this representation of facial surfaces is depth cue invariant—that the representation from an adaptor defined by one depth cue transfers to a test stimulus defined by another depth cue. We found that such 3-D face representations are depth cue invariant. Given that different depth cues appear to be processed by different visual areas across parallel visual
pathways, an integrative process would be needed to yield a single depth cue invariant representation, putatively in the ventral stream. The results have important implications for cortical and computational models of object recognition and highlight the importance of considering depth as a cue to complex object representations.

**Keywords:** face representation, depth cue invariance, object recognition, face identity aftereffect, antiface, shading, texture, structure from motion, stereo disparity

### Acknowledgments

We are grateful to Singular Inversions (provider of FaceGen Modeller) for providing us with a custom software tool and instructions for generating 3-D antiface models.

Commercial relationships: none.
Corresponding author: Reza Farivar.
Email: reza.farivar@mcgill.ca.
Address: McGill Vision Research Unit, Montreal General Hospital, Montreal, Québec, Canada.

### References


Kobatake, E., & Tanaka, K. (1994). Neuronal sele-
tivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71, 856–867.


