Université de Montréal

L'impact de la stéréoscopie dans la reconnaissance, la perception et la constance de forme 3D

par

Mercédès Aubin

Département de psychologie
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en Psychologie - option Sciences cognitives et neuropsychologie

17 avril 2014

Résumé

Les buts des recherches présentées dans cette thèse étaient d'évaluer le rôle de la stéréoscopie dans la reconnaissance de forme, dans la perception du relief et dans la constance de forme.

La première étude a examiné le rôle de la stéréoscopie dans la perception des formes visuelles en utilisant une tâche de reconnaissance de formes. Les stimuli pouvaient être présentés en 2D, avec disparité normale (3D) ou avec disparité inversée. La performance de reconnaissance était meilleure avec les modes de présentation 2D et 3D qu'avec la 3D inversée. Cela indique que la stéréoscopie contribue à la reconnaissance de forme.

La deuxième étude s'est intéressée à la contribution conjointe de l'ombrage et de la stéréoscopie dans la perception du relief des formes. Les stimuli étaient des images d'une forme 3D convexe synthétique présentée sous un point de vue menant à une ambigüité quant à sa convexité. L'illumination pouvait provenir du haut ou du bas et de la gauche ou de la droite, et les stimuli étaient présentés dichoptiquement avec soit de la disparité binoculaire normale, de la disparité inversée ou sans disparité entre les vues. Les participants ont répondu que les formes étaient convexes plus souvent lorsque la lumière provenait du haut que du bas, plus souvent avec la disparité normale qu'en 2D, et plus souvent avec absence de disparité qu'avec disparité inversée. Les effets de direction d'illumination et du mode de présentation étaient additifs, c'est-à-dire qu'ils n'interagissaient pas. Cela indique que l'ombrage et la stéréoscopie contribuent indépendamment à la perception du relief des formes.

La troisième étude a évalué la contribution de la stéréoscopie à la constance de forme, et son interaction avec l'expertise perceptuelle. Elle a utilisé trois tâches de discrimination séquentielle de trombones tordus ayant subi des rotations en profondeur. Les stimuli pouvaient être présentés sans stéréoscopie, avec stéréoscopie normale ou avec stéréoscopie inversée. Dans la première moitié de l’Exp. 1, dans laquelle les variations du mode de présentation étaient intra-sujets, les performances étaient meilleures en 3D qu’en 2D et qu’en 3D inversée. Ces effets ont été renversés dans la seconde moitié de l’expérience, et les coûts de rotation sont devenus plus faibles pour la 2D et la 3D inversée que pour la 3D. Dans les Exps. 2 (variations intra-sujets du mode de présentation, avec un changement de stimuli au milieu de l’expérience) et 3 (variations inter-sujets du mode de présentation), les effets de rotation étaient en tout temps plus faibles avec stéréoscopie qu’avec stéréoscopie inversée et qu’en 2D, et plus faibles avec stéréoscopie inversée que sans stéréoscopie. Ces résultats indiquent que la stéréoscopie contribue à la constance de forme. Toutefois, cela demande qu’elle soit valide avec un niveau minimal de consistance, sinon elle devient stratégiquement ignorée.

En bref, les trois études présentées dans cette thèse ont permis de montrer que la stéréoscopie contribue à la reconnaissance de forme, à la perception du relief et à la constance de forme. De plus, l’ombrage et la stéréoscopie sont intégrés linéairement.

Mots clés : Reconnaissance de forme, constance de forme, stéréoscopie, perception du relief

Abstract

The goals of the researches presented in this thesis were to evaluate the role of stereopsis in shape recognition, in relief perception, and in shape constancy.

The first study examined the role of stereopsis in visual shape perception using a recognition task. The stimuli were presented with null binocular disparity (i.e. 2D), normal binocular disparity (3D) or reversed disparity. Recognition performance was better with 2D and 3D displays than with reversed 3D. This indicates that stereopsis contributes to shape recognition.

The second study examined the joint contribution of shading and stereopsis to the relief perception of shape. The stimuli were the images of a synthetic convex 3D shape seen from viewpoints leading to ambiguity as to its convexity. Illumination either came from above, or below and from the right or the left, and stimuli were presented dichoptically with either normal binocular disparity, reversed disparity, or no disparity between the views presented at each eye. Participants responded “convex” more often when the lighting came from above than from below. Also, participants responded that the shape was convex more often with normal than with zero disparity, and more often with 2D than with reversed stereopsis. The effects of lighting direction and display mode were additive; i.e. they did not interact. This indicates that shading and stereopsis contribute independently to shape perception.

The third study assessed the contribution of stereopsis to shape constancy and how it interacts with perceptual expertise using three sequential matching tasks with bent paperclips rotated in depth. Stimuli were presented without stereopsis, or with normal or reversed stereopsis. In the first half of Exp. 1, where display mode variations were within-subject, the performances were better with stereoscopic displays than with 2D or reversed stereoscopic presentations. In the second half of the experiment, the rotation costs became weaker for the 2D and reversed 3D display modes than for the 3D one. In Exps. 2 (display mode within-subject, with stimuli switched halfway into the experiment) and 3, (display mode between-subjects) the rotation effect was consistently weaker with normal stereo than with either 2D or reversed stereoscopic displays. These experiments also demonstrate an advantage of reversed stereo over 2D presentations. This indicates that stereo may contribute to shape constancy. This, however, requires stereoscopic information to be valid with a minimal degree of consistency. Otherwise, stereo may become strategically ignored.

In a nutshell, the three studies presented in this thesis showed that stereo contributes to shape recognition, relief perception and shape constancy. Furthermore stereopsis and shading are integrated independently.

Keywords: Shape recognition, shape constancy, stereopsis, relief perception

Table des matières

Cadre théorique

Liste des tableaux

Article 3

Listes des figures

Article 1

Article 2

Article 3

Liste des abréviations

2D: deux dimensions

3D trois dimensions

ER: Error rate

Exp: Expérience (experiment)

IRMf : Imagerie par résonance magnétique fonctionnelle

IT : Cortex inféro-temporal

PNA: Propriété non-accidentelle

RT: Response time

Pour Pipit et Anne, et à la mémoire de Miguel

Remerciements

Je tiens à remercier mon directeur de recherche, Martin Arguin, pour son support et son aide, ainsi que Frédéric Gosselin et Pierre Jolicoeur pour leurs commentaires.

Introduction

Dans l’Odyssée d’Homère, Ulysse se retrouva au pays des Cyclopes, coincé dans l’antre de Polyphème, un redoutable anthropophage ne disposant que d’un seul œil. Usant de ruse, Ulysse enivra Polyphème et profita de son sommeil pour lui crever son unique œil. Si nous étions, comme Polyphème, des cyclopes, nous n’aurions qu’un œil, et donc pas de vision binoculaire, ce qui nous priverait de l’information de profondeur stéréoscopique. Dans quelle mesure cela nous affecterait-il? Nous pouvons nous demander quelle est l’influence de la stéréoscopie dans la perception et la représentation des formes, et la perception du relief. Les différents articles présentés dans cette thèse tentent de répondre à ces questions.

Le premier a pour but d’évaluer l’impact de la stéréoscopie dans la reconnaissance de forme. Le second a pour but d’étudier l’intégration de la stéréoscopie et de l’ombrage dans la perception du relief de formes. Enfin, le troisième article cherche à évaluer le rôle de la stéréo dans la constance de forme. Ces articles sont précédés d’une recension des écrits portant sur les théories de la représentation des formes, sur l’impact des informations de profondeur et sur les bases cérébrales qui sous-tendent la perception et la représentation des formes.

Cadre théorique

Théories de la représentation des formes visuelles

Marr et Nishihara (1978) ont proposé un modèle structural pour la représentation des objets tridimensionnels. Celui-ci postule une représentation stable et invariante au changement de l’image rétinienne qui est obtenue en identifiant les différentes parties des objets et leurs relations. S’inspirant de ce modèle, Biederman (1987) a élaboré une théorie structurale de la représentation des objets, la théorie de la reconnaissance par composantes (RPC), selon laquelle la perception des formes se fait par l'extraction de propriétés particulières des contours appelées propriétés non-accidentelles (PNA), soit la cotermination, la symétrie, la courbure, le parallélisme et la collinéarité. Les caractéristiques de l'image rétinienne (qui est bidimensionnelle, i.e. 2D) d'un objet quant à ces propriétés signalent de manière fiable (exception faite des points de vue dits accidentels) et invariante des propriétés de la forme de l'objet tridimensionnel (3D). Les composantes constituant l'objet, appelées géons (pour ions géométriques) sont déterminées par la combinaison particulière de leurs PNA. La détection des PNA est résistante à la rotation en profondeur (i.e. 3D), ce qui rend la perception des géons invariante au point de vue. Une fois les géons encodés, les relations spatiales entre les composantes sont déterminées. Ce traitement donne naissance à une représentation structurale 3D d'objets complexes (i.e. constitués de plus d'une partie) qui est invariante à l'orientation. Ainsi, un objet sera reconnu sous n'importe quel point de vue 3D du moment que les PNA sont visibles. Biederman et Gerhardstein (1993) ont obtenu des résultats appuyant cette hypothèse d'invariance au point de vue. Toutefois, certaines conditions s'imposent pour que cette invariance s'applique. D'abord, les objets doivent posséder une description géonique structurale (traduction par l'auteure de «geon structural description»), c'est-à-dire qu'il doit être possible de les décrire par leurs géons et les relations spatiales entre ceux-ci. Ensuite, ces descriptions doivent être distinctes entre les stimuli et elles doivent se maintenir à travers différents points de vue (Biederman & Gerhardstein, 1993).

Dans le même ordre d'idée, Hummel et Biederman (1992) ont créé un modèle de réseau neuronal qui parvient à générer des descriptions structurales des objets. Celui-ci présente une invariance à la rotation en profondeur (i.e. 3D) mais il est sensible à la rotation dans le plan (i.e. 2D). Comme le propose Biederman (1987), le modèle extrait les différentes composantes et les relations spatiales entre elles. Ces relations sont codées sous des termes tels: en haut de, en bas de, à côté de, etc., ce qui les rend donc sensibles à la rotation 2D mais pas à la rotation 3D.

Biederman et Bar (1999) ont démontré l'importance des PNA dans l'invariance à la rotation 3D proposée dans les modèles de Biederman (1987) et Hummel et Biederman (1992). Les effets de rotation sur les temps de réponse et les taux d'erreurs dans une tâche d'appariement séquentiel sont très faibles lorsque les objets se distinguent par des différences de PNA, alors que ces effets sont beaucoup plus marqués lorsque les objets se distinguent par des différences de propriétés métriques (e.g. longueur, largeur, degré de courbure, angle d'attache, etc.; Biederman & Bar). Cette plus grande sensibilité aux PNA qu'aux propriétés métriques est aussi présente dans les neurones du cortex inféro-temporal (IT) chez le macaque (Vogels, Biederman, Bar & Lorincz, 2001). Vogels et ses collègues ont mesuré l'activité de neurones individuels de l'aire IT lors de la présentation de stimuli et ont montré que ceux-ci étaient beaucoup plus sensibles aux PNA qu'aux propriétés métriques. En effet, la différence d'activation produite par la rotation d'une forme donnée est d'une amplitude semblable à celle produite par le changement d'une propriété métrique, mais de moins grande amplitude que la modification d'une PNA. Par ailleurs, Gibson et al. (2007) ont montré, avec la technique des bulles, que la PNA de cotermination était utilisée par les pigeons comme par les humains pour reconnaître les formes. Ce résultat indique que malgré les différences dans l'anatomie du système visuel des pigeons et des humains, les deux espèces présentent un biais pour les propriétés non accidentelles.

Dans la même optique, Pizlo et Stevenson (1999) ont cherché à savoir quelles sont les propriétés visuelles dont dépend la constance de forme. Plutôt que de s'attarder aux propriétés non-accidentelles, ils ont vérifié d'autres caractéristiques à l'aide d'une tâche de discrimination. Les performances étaient meilleures lorsque les stimuli étaient des polyhèdres dont les contours étaient planaires (i.e. chaque face dans un seul plan), qu'ils étaient symétriques et qu'il n'y avait qu'un nombre restreint d'interprétations possibles de la forme 3D à partir des contours, que lorsque ces différentes propriétés étaient absentes (Pizlo & Stevenson).

Contrairement aux propositions de la théorie de Biederman, certaines études ont rapporté une détérioration des performances suite à une rotation 3D (Hayward & Tarr, 1997; Tarr, Williams, Hayward & Gauthier 1998) même si les stimuli utilisés respectaient les conditions d'invariance proposées par Biederman et Gerharstein (1993). Les théories de la représentation basée sur les vues permettent d'expliquer ces effets (Hayward & Tarr). Elles proposent qu'ils proviennent de procédures de normalisation (Tarr, 1995). Selon ces théories, les objets sont représentés par les coordonnées des traits de leurs images 2D, et des processus comme l'interpolation (Edelman & Bülthoff, 1992; Poggio & Edelman, 1990), la combinaison linéaire (Ullman & Basri, 1991), la rotation mentale (Shepard & Metzler, 1971; Tarr & Pinker, 1989) et l'alignement des descriptions picturales (Ullman, 1989) permettent à l'image entrante d'un objet connu observé sous une orientation 2D ou 3D non-familière d'être appariée à des images de cet objet encodées en mémoire dans des orientations familières. Ces approches prédisent que plus la distance angulaire entre les images est grande, plus les processus de normalisation seront coûteux. Les résultats de Leek (1998) montrent des effets de rotations dans une tâche de vérification mot-image lorsque les stimuli sont des objets mono-orientés (habituellement visionnés sous un seul point de vue), mais pas lorsqu'ils sont poly-orientés (habituellement visionnés sous plusieurs points de vue.) Ceux-ci appuient l'idée que les formes sont encodées avec des représentations spécifiques au point de vue. Notons par contre que les résultats de Gauthier et ses collaborateurs (2002) ont montré que les substrats neuronaux impliqués dans la réalisation d'une tâche de rotation mentale (sans reconnaissance d'objet requise) étaient en partie différents des substrats neuronaux impliqués dans les tâches de reconnaissance d'objets désorientés. Ceci indique que la rotation mentale n'est probablement pas le processus de normalisation utilisé dans les tâches de reconnaissance de forme. D'ailleurs, des résultats comportementaux appuient ce dernier point puisque les patrons de résultats pour des stimuli identiques diffèrent selon que la tâche en soit une de rotation mentale ou de reconnaissance d'objets (Hayward, Zhou, Gauthier & Harris, 2006). En effet, les temps de réponse augmentent de manière linéaire lors de la tâche de rotation mentale, mais sont mieux décrits par une fonction quadratique pour la tâche de reconnaissance. Qu'importe le type de processus de normalisation utilisé, un fait bien établi est que les coûts de rotation diminuent avec l'entraînement (Jolicoeur, 1985).

Comme le montrent les paragraphes précédents, il n'y a pas de consensus entre invariance et dépendance au point de vue des représentations visuelles. Les travaux de Burgund et Marsolek (2000) laissent croire que la représentation des formes n'est pas nécessairement qu'invariante ou dépendante à l'orientation. En effet, ils suggèrent la présence de deux réseaux neuronaux impliqués dans la représentation des formes. L'un entrainerait des performances invariantes à l'orientation et l'autre l'inverse. Dans une tâche d'amorçage, ils ont montré que l'amorçage était dépendant de l'orientation lorsque les stimuli étaient présentés à l'hémisphère droit, et invariant lorsque les stimuli étaient présentés à l'hémisphère gauche. Ces résultats suggèrent qu'il y a un système plus sensible aux détails des images qui traiterait les formes en se basant sur les vues dans l'hémisphère droit, et qu'il y a un système de traitement qui représente les formes d'une manière invariante au point de vue dans l'hémisphère gauche (Burgund & Marsolek, 2000).

Les théories basées sur les vues ont surtout utilisé les coûts de rotation pour appuyer leurs propositions. Toutefois, Stankiewicz (2002) a montré qu'il était possible qu'il y ait des coûts de rotation dans la représentation des formes malgré une indépendance au point de vue. En effet, à l'aide d'un paradigme de masquage par le bruit, il a révélé que l'orientation était encodée par le système qui traite les formes, mais indépendamment des informations sur la forme. Cela correspond davantage aux théories structurales qu'aux théories des vues, mais sans pour autant garantir l'invariance à l'orientation. En effet, l'information provenant du point de vue peut-être informative, et si le système visuel utilise l'information provenant de la forme et celle relative au point de vue, il s'ensuit que les performances seront meilleures lorsque les objets sont observés sous des points de vue familiers que des points de vue nouveaux. Notons aussi qu'indépendamment des processus visuels impliqués dans la représentation des formes, il y a une certaine dépendance au point de vue qui est inhérente à la tâche effectuée et aux stimuli utilisés (Tjan & Legge, 1998). Des simulations faisant usage d'un observateur idéal montrent que l'utilisation de stimuli qui sont des objets géométriques simples (e.g. cylindre, cône) entraîne une bien moins grande dépendance au point de vue que des stimuli en forme de fève ou de câble. Pour ce qui est des objets complexes, si certaines parties entraînent une faible dépendance au point de vue et que d'autres entraînent une forte dépendance, l'objet complexe dans sa totalité entraînera une forte dépendance au point de vue (Tjan & Legge, 1998).

L'impact des informations de profondeur

Les informations de profondeur utilisées dans la représentation des formes ne sont pas les mêmes selon les théories structurales et celles basées sur les vues. En effet, Biederman (1987) propose une information 3D implicite qui ne repose que sur le traitement des contours alors que d'autres types d'informations de profondeur, tels le gradient de texture, l'ombrage ou la stéréoscopie, ne seraient pas utilisés. Les théories basées sur les vues admettent la possibilité d'une contribution des indices de profondeur à la représentation de la forme mais demeurent ambigües quant à la présence effective d'une telle contribution. En effet, puisque les informations de profondeur comme l'ombrage et la texture font partie des images 2D des objets, il est possible que celles-ci soient utilisées par les processus de normalisation.

L'hypothèse selon laquelle l'information de profondeur ne contribue pas à la perception est appuyée par les résultats de Pizlo et ses collègues (Pizlo, 2008; Pizlo, Li & Francis, 2005; Chan, Stevenson, Li & Pizlo, 2006; Pizlo, Li, Steinman, 2008) qui se sont intéressés au rôle de l'information de profondeur stéréoscopique dans la représentation des formes. Ils suggèrent que la représentation des formes et la constance des formes se font à partir de contraintes de simplicité monoculaires à priori (planarité de contours de surface, compacité et symétrie) et que l'information de profondeur binoculaire n'est que très secondaire, et ni suffisante, ni nécessaire à la représentation des formes (Pizlo, 2008). D'abord, Pizlo et ses collègues (2005) ont montré avec une tâche de détermination de l'amplitude de mouvement de formes présentées stéréoscopiquement qui mettait en conflit la disparité binoculaire et les contraintes de simplicité, que la disparité binoculaire était ignorée par les participants. Ensuite, Chan et ses collaborateurs ont testé des participants lors de tâches de discrimination séquentielle d'objets et ont comparé les résultats de ces derniers aux résultats de différents modèles de reconstruction de formes. Les performances des modèles de reconstruction monoculaire corrèlent avec les performances monoculaires et binoculaires humaines, mais ce n'est pas le cas des modèles de reconstruction binoculaire. Cela suggère que le système visuel humain utiliserait des mécanismes basés sur des contraintes monoculaires.

Pour ce qui est du rôle des informations de profondeur monoculaires, Biederman et Ju (1988), ont montré que la dénomination d'objets est aussi rapide si les stimuli sont des contours d'objets que s'ils sont des photographies. De plus, le modèle structural de Pentland (1989 – très proche de celui de Biederman, s'en distinguant principalement par le type de traits volumétriques postulés) parvient à recouvrer les surfaces des formes uniquement à partir de l'information provenant des contours. Ces résultats indiquent que l'information de profondeur provenant des contours est suffisante à une bonne représentation des objets. Biederman et Ju et Pentland n'ont toutefois pas vérifié directement l'impact de l'information de profondeur sur l'invariance à l'orientation.

Le rôle de l'information binoculaire. Les paragraphes précédents suggèrent l'absence de rôle de l'information de profondeur dans la représentation des formes. Pourtant, pour ce qui est de l'information stéréoscopique, certaines études montrent le contraire. D'abord, Julesz (1960, 1971) a montré qu'il était possible de reconnaître des formes 3D (paraboloïdes) présentées avec des stéréogrammes de points aléatoires, qui sont dépourvus d'information de profondeur monoculaire disponible et où seule la disparité binoculaire y définit les formes visibles. Des études plus récentes ont démontré qu'il était possible de reconnaître des surfaces 3D courbées uniquement définies par des stéréogrammes de points aléatoires (De Vries, Kappers & Koenderink, 1993; Uttal, Davis & Welke, 1994; Uttal, Davis, Welke & Kakarala, 1988; Vreven, 2006.) Par ailleurs, Burke (2005) a démontré, avec des tâches de discrimination et de reconnaissance de formes, qu'il est possible de réduire l'effet de rotation 3D en ajoutant une information stéréoscopique. Burke, Taubert et Higman (2007) ont rapporté des effets similaires sur la discrimination de visages désorientés. Bennett et Vuong (2006) ont montré que cette réduction de l'effet de rotation se maintenait pour différents niveaux de difficulté des tâches et différentes amplitudes de rotation. Bien que les effets de rotation soient réduits par l'ajout d'information binoculaire dans les expériences de Burke et de Bennett et Vuong, ils ne disparaissent jamais complètement. Il est intéressant de noter que pour ces dernières études de la stéréoscopie a contribué à la constance de forme malgré le fait que l'information de profondeur monoculaire était riche; les stimuli étaient des photographies pour Burke (2005) et Burke et al. (2007) et contenaient de l'ombrage et de l'occlusion pour Bennett et Vuong.

Cet impact de la stéréoscopie va à l'encontre de la théorie de RPC et des modèles de Pizlo et ses collègues (Pizlo, 2008; Pizlo & al, 2005; Chan & al, 2006; Pizlo & al, 2008). Toutefois, les stimuli utilisés dans ces trois études (Bennett & Vuong, 2006; Burke, 2005; Burke & al, 2007) ne répondaient pas aux conditions d'invariance au point de vue de Biederman et Gerharstein (1993), ni aux contraintes de simplicité proposées par Pizlo et ses collègues (e.g. stimuli en forme de câbles tordus ou d'amibes). Il est aussi intéressant de noter que de plus récents résultats de Li et Pizlo (2011) ont montré que la stéréoscopie pouvait effectivement contribuer à la constance de formes respectant les contraintes de simplicité des contours. De leur côté, Pasqualotto et Hayward (2009) ont démontré, avec une tâche de discrimination de formes familières très différenciables désorientées, que bien que l'information binoculaire puisse permettre de recouvrer des informations fines et précises, l'extra-spécificité entrainée par la stéréoscopie diminuerait l'importance accordée aux similarités plus générales entre les éléments picturaux des images et pourrait ainsi nuire aux performances de discrimination sous certaines conditions particulières.

Quelques études un peu plus anciennes se sont intéressées au rôle de la stéréoscopie, mais en évaluant l'impact qu'avait la stéréoscopie inversée dans la perception visuelle. D'abord, Wheatstone (1852), a utilisé un pseudoscope (stéréoscope qui présente à l'œil droit l'image qui irait normalement à l'œil gauche et vice-versa) et a ressenti que les formes vues à travers cet instrument avaient la profondeur renversée. Plus récemment, des expériences utilisant des lunettes à prismes qui inversent la stéréoscopie ont été menées (Shimojo & Nakajama, 1981; Ichikawa, Egusa, Nakatsuka, Amano, Ueoa & Tashiro, 2003; Ichikawa & Egusa, 1993; Yellott & Kaiwi, 1979). Elles ont montré qu'à long terme, l'adaptation à la stéréoscopie inversée était possible. Toutefois, les expériences portaient sur l'évaluation de la profondeur et non sur l'effet direct de l'inversion sur la reconnaissance et la constance de formes. D'ailleurs, au meilleur de notre connaissance, aucune étude ne s'est intéressée à ce dernier point. Il est tout de même intéressant de noter que bien que l'adaptation à la stéréoscopie inversée entraînait chez la plupart des sujets une inversion de la profondeur lorsque les lunettes étaient retirées, elle faisait aussi en sorte qu'il y avait un changement dans le poids des différentes informations de profondeur monoculaire avec l'adaptation. En effet, la disparité binoculaire finissait par être ignorée et l'occlusion et la perspective linéaire avaient davantage d'influence sur les jugements de profondeur qu'avant l'adaptation (Ichikawa & Egusa, 1993).

Le rôle des informations monoculaires.

Des informations de profondeur monoculaires comme la texture, l'ombrage et les réflexions spéculaires ont aussi un impact dans la perception de l'inclinaison des surfaces, dans l'évaluation locale de surface et dans la constance de forme (Saunders & Backus, 2006; Blais, Arguin & Marleau, 2009; Fleming, Torralba & Andelson, 2004; Khang, Koenderinck & Kappers, 2007; Mingolla & Todd, 1986; Ramachandran, 1988; Todd, Norman, Koenderinck & Kappers, 1997). Nous nous attarderons à l'ombrage dans le cadre de cette thèse. C'est une information qui permet d'estimer les formes 3D (Mingolla & Todd, 1986; Ramachandran, 1988). Toutefois, Erens, Kappers et Koenderink (1993) ont démontré que bien que cet indice permette de déterminer la convexité ou la concavité d'ellipses, il ne permet pas à lui seul de différencier les formes elliptiques et hyperboliques. De plus, Bülthoff et Mallot (1988) ont montré que toutes les informations de profondeur n'ont pas le même poids. En effet, l'ombrage présenté avec disparité binoculaire ou l'ombrage seul permettent une moins bonne estimation de la profondeur locale des surfaces que les contours présentés avec disparité binoculaire. Cependant, plus il y a d'information disponible et plus l'estimation de la profondeur s'améliore (Bütlhoff & Mallot).

L'interaction des informations binoculaire et monoculaires. Norman, Todd et Orban (2004) ont mesuré l'interaction entre des informations monoculaires (réflexions spéculaires, texture, contours, ombrage) et stéréoscopique dans la discrimination de blobs et ont montré que la stéréoscopie n'améliorait les performances de discrimination que lorsque les stimuli avaient comme information de profondeur disponible de la texture et des contours, de l'ombrage et des contours, ou de la texture de l'ombrage et des contours, mais pas lorsque leurs informations disponibles étaient des réflexions spéculaires, des réflexions spéculaires et des contours ou des réflexions spéculaires, de l'ombrage et des contours. Saunders et Backus (2006) ont montré que la stéréoscopie permettait de réduire le seuil de discrimination d'inclinaisons de surfaces définies uniquement par la texture. Toutefois, il y avait des variations entre les participants, et pour certains, l'information binoculaire n'avait aucun poids. Pour ce qui est de la perception du relief, la disparité binoculaire semble influencer l'impact de la texture (Adams & Mamassian, 2004). En effet, alors qu'avec de la texture seule, des formes ambigües étaient perçues comme convexes, l'ajout de stéréoscopie indiquant l'inverse rendait la perception concave. Cela indique que ces deux informations sont combinées de manière non linéaire. De plus, l'ombrage influence l'impact de la stéréoscopie dans la perception de surfaces sinusoïdales (Wright & Ledgeway, 2004), ce qui indique aussi une combinaison non linéaire de ces deux informations. Par contre, Johnston, Cumming et Parker (1993) ont montré, avec une tâche de jugement de forme (déterminer l'étendue en profondeur de cylindres) que la texture et la stéréoscopie étaient intégrées indépendamment l'une de l'autre, et Doorschot, Kappers et Koenderink (2001) ont montré qu'il en était de même pour l'ombrage et la stéréoscopie dans l'évaluation locale de la surface de formes.

Stevens, Lees et Brookes (1991) ont évalué l'impact de l'incohérence entre l'information stéréoscopique et l'information provenant des contours sur la description qualitative de surfaces (i.e. la surface décrite par l'information stéréoscopique était différente de celle décrite par les contours). Il y a eu beaucoup de différences entre les sujets dans les descriptions. En effet, certains ignoraient totalement l'information stéréoscopique et ne comptaient que sur les contours alors que d'autres se fiaient d'avantage à elle. Certains tentaient, malgré l'incohérence, de combiner les deux pour en arriver à des représentations très instables et ressentaient beaucoup d'incertitude quant à leurs descriptions (Stevens & al, 1991). Norman et Todd (1995) ont évalué la perception de surfaces 3D lorsque l'information stéréoscopique et une information de profondeur créée par le mouvement étaient inconsistantes. Ils ont trouvé que dans les cas de conflit, une des deux informations était supprimée en fonction de l'orientation des courbures. Du moment où la direction de courbure signalée par l'une des deux informations était valide, l'autre information était ignorée (Norman & Todd). Lorsqu'il y a inconsistance entre l'information stéréoscopique et l'information donnée par le gradient de texture, l'évaluation de l'inclinaison en profondeur de surfaces utilise davantage la texture que la stéréoscopie (Saunders & Backus, 2006). Cela dépend toutefois du type de texture. En effet, les textures qui sont symétriques et dont l'alignement correspond à l'axe d'inclinaison ont plus de poids que celles dont l'alignement et perpendiculaire à l'axe ou celles qui sont isotropiques (Saunders & Backus). Dans le même ordre d'idée, Braunstein (1986) et Stevens et Brookes (1988) ont montré que lorsque la stéréoscopie et les informations monoculaires (occlusion et gradient de texture) étaient inconsistantes, l'information monoculaire outrepassait l'information binoculaire.

Les a priori. Bien qu'il soit possible de percevoir des formes 3D définies par l'ombrage (Ramachandran, 1988), on les perçoit avec la contrainte qu'elles soient éclairées avec une seule source d'illumination, et l'ombrage peut mener à différentes interprétations possibles en fonction de la localisation de la source. L'a priori d'illumination par le haut est un biais perceptif qui permet de résoudre ces ambiguïtés (Adams, 2007; Connor, 2001; Gerardin, Montalembert & Mamassian, 2007; Kleffner & Ramachandran, 1992). En effet, il est postulé lors du traitement perceptif que la lumière vient du haut, et une lumière provenant du bas entraine un renversement de la profondeur perçue. De plus, il y aurait un biais pour la lumière provenant de la gauche (Mamassian & Goutcher, 2001; Sun & Perona, 1998), mais cela ne fait pas consensus. En effet, McManus, Buckman et Woolley (2004) ont montré qu'il n'y avait pas de différence dans le jugement de convexité de formes selon que la lumière vienne de la gauche ou de la droite lorsque des stimuli étaient présentés pour une durée indéfinie, mais le biais de la lumière provenant de la gauche se manifestait lorsque les stimuli étaient présentés brièvement (entre 200 et 1000 ms). Mamassian et Landy (2001) ont montré que lorsque les a priori d'illumination par le haut et de point de vue d'observation par le haut (interprétation de la forme comme si le point de vue d'observation était par le haut) menaient à des interprétations différentes, l'a priori qui l'emportait était celui pour lequel l'information correspondante (l'ombrage pour l'a priori d'illumination du haut, et les contours pour l'a priori du point de vue d'observation par le haut) était la plus fiable. Les auteurs ont modulé la fiabilité de ces a priori en faisant varier le contraste de l'ombrage ou des contours. Ce phénomène est analogue a celui observé pour l'intégration d'informations de profondeur incongruentes, où la source la plus fiable a le plus de poids et peut même dominer entièrement l'autre (Norman & Todd, 1995; Saunders & Backus, 2006).

Différences individuelles

Un point important à aborder concernant les études comportementales sur la représentation des formes est la présence de différences individuelles. Rappelons que Stevens, Lees et Brooke (1991) et Saunders et Backus (2006) ont trouvé des différences individuelles dans l'utilisation de diverses informations de profondeur. Stone, Buckley et Moger (2000) ont quant à eux tenté de décrire des profils individuels de reconnaissance d'objets. Chez les observateurs qui sont rapides dans l'apprentissage de la reconnaissance d'objets, il y a un profil typique : il y a une préférence pour l'utilisation des contours externes pour faire des jugements de profondeur, et les observateurs présentant ce profil utilisent davantage les contours et la texture que le mouvement pour les jugements de profondeur lorsqu'il y a incongruence entre les indices de profondeur.

Substrats neuronaux du traitement des formes et informations de profondeur

Il y a une dissociation ventrale et dorsale dans les substrats neuronaux impliqués dans la perception visuelle. D'un côté, la voie dorsale s'occupe de traiter la localisation spatiale et de la guidance visuelle de l'action, et de l'autre, la voie ventrale est impliquée dans la reconnaissance d'objets (Goodale & Milner, 1992; Goodale et Westwood 2004; Ungerleider & Mishkin, 1982). Dans la voie ventrale, certains neurones de l'aire IT du macaque présentent une invariance à l'orientation 3D pour des objets connus (Booth & Rolls, 1998). En effet, bien que lors d'une tâche de fixation, la plupart des neurones dont l'activité était mesurée dans l'aire IT répondaient sélectivement aux objets familiers présentés sous des points de vue particuliers, quelques-uns avaient des réponses sélectives à des objets spécifiques, peu importe l'orientation sous laquelle ils étaient présentés. Il semble donc y avoir deux modes de représentation dans le cortex IT, un similaire à celui proposé par les théories des vues et l'autre compatible avec les théories structurales. Perrett et al. (1991) ont trouvé des résultats semblables en présentant des têtes sous différents points de vue à des macaques. La plupart des neurones du sulcus temporal supérieur dont l'activité a été mesurée répondaient sélectivement au point de vue, mais certains répondaient aux têtes d'une manière invariante à l'orientation. Yamane, Carlson, Bowman, Wang et Connor (2008) ont trouvé des résultats similaires avec une méthode élégante consistant à présenter des stimuli aléatoires à des neurones, puis les faire évoluer pour que ceux-ci correspondent davantage aux préférences des neurones du cortex IT. Ils ont montré qu'une importante partie des neurones répondaient sélectivement à des formes 3D et à l'organisation spatiale des fragments des formes, ce qui est en accord avec les théories structurales. Toutefois, le cadre de référence était centré en fonction de l'objet, mais ne tournait pas avec l'objet lorsque ce dernier subissait des rotations, ce qui est consistant avec les théories des vues.

Toujours dans la voix ventrale, les substrats neuronaux impliqués dans le traitement de la géométrie des objets et ceux qui sont impliqués dans le traitement des informations de surface seraient différents. En effet, Cavina-Pratesi, Kentridge, Heywood et Milner (2010) ont montré, dans une tâche consistant à discriminer des formes ou des textures, que l'attention aux attributs géométriques élicitait une activation du complexe occipital latéral, alors que l'attention à la texture suscitait des activations du sulcus collatéral. Aussi, les auteurs ont trouvé une double-dissociation chez deux sujets agnosiques. Un premier, dont le complexe occipital latéral ne s'activait pas, ne parvenait pas à discriminer les formes, mais discriminait les textures. Un autre patient, dont le sulcus collatéral ne s'activait pas, présente la dissociation opposée. De plus, il semble que le complexe occipital latéral est sensible aux formes globales, et non aux propriétés visuelles élémentaires qui les constituent (i.e. contours, information de surface, etc.) En effet, Kourtzi et Kanwisher (2001) ont montré, avec de l'imagerie par résonance magnétique fonctionnelle (IRMf), qu'il y avait adaptation du complexe latéral occipital lorsqu'une forme était présentée deux fois d'affilée avec des contours externes différents d'une fois à l'autre. Par contre, il n'y avait pas d'adaptation lors de la présentation séquentielle de formes différentes avec les mêmes contours externes.

Des études d'IRMf montrent que les substrats neuronaux impliqués dans le traitement des formes visuelles utilisent les différentes informations de profondeur et les combinent pour en arriver à des représentations efficaces. Par exemple, Welchman, Deubelius, Conrad, Bülthoff et Kourtzi (2005) ont étudié les corrélats neuronaux de la combinaison de disparité binoculaire et de perspective dans la représentation de formes 3D. Ils ont fait varier ces deux informations de profondeur dans une tâche consistant à évaluer l'angle d'attache entre deux rectangles. Leurs résultats suggèrent que les aires primaires du cortex visuel (V1, V2, V3 et V4) sont sensibles au type d'information de profondeur, alors que les aires extra-striées V5 et le cortex occipital latéral sont plutôt sensibles à la forme 3D créée par la combinaison des deux types d'information.

On a aussi étudié comment les régions cérébrales du macaque impliquées dans la représentation des formes utilisaient les informations de profondeurs. Liu, Vogels et Orban (2004) ont découvert que des neurones du cortex inféro-temporal (IT) étaient sélectifs à l'inclinaison de surfaces, lorsque cette information provenait de la texture, peu importe le type (cercles, quadrillage, ligne), et que cette sélectivité était fortement corrélée à la sélectivité à l'inclinaison amenée par la disparité binoculaire. Cette invariance au type d'indice laisse supposer que ces neurones codent de l'information quant aux formes 3D (Liu & al., 2004).

Il est intéressant de souligner que l'activité de V3 et de hMT+/V5, qui font partie de la voie dorsale, ainsi que l'activité du complexe occipital latéral, qui fait partie de la voie ventrale, corrèlent avec les performances dans une tâche de jugement de symétrie de formes définies par des stéréogrammes à points aléatoires (Chandrasekaran, Canon, Dahmen, Kourtzi & Welchman, 2007). Ceci suggère que la voie dorsale est aussi impliquée dans la perception des formes. De plus, la perception de formes définies par la texture entraîne l'activation de régions du sulcus intra-pariétal dans la voie dorsale, et celle de régions du gyrus inféro-temporal, dans la voie ventrale (Georgieva, Todd, Peeters & Orban, 2008.) Par contre, la perception de formes définies uniquement par l'ombrage entraîne de l'activation du gyrus inféro-temporal, mais pas d'activation du sulcus intra pariétal (Georgieva et al., 2008). Cela indique que le gyrus inféro-temporal est important pour l'extraction de l'information sur la forme 3D amenée par diverses informations de profondeur. Par contre pour ce qui est de l'analyse des formes 3D dans la voie dorsale, l'ombrage serait moins important que d'autres sources d'information (Georgieva et al., 2008). On a aussi trouvé que 50% des neurones de l'aire IT du macaque étaient sensibles non seulement à la forme, mais aussi à la disparité binoculaire, ce qui suggère que IT est impliqué dans la reconstruction de surfaces 3D (Uka, Tanaka, Yoshiyama, Kato & Fujita, 2000). De plus, chez le macaque, dans la voie ventrale (V4), on a trouvé des neurones sensibles à la disparité binoculaire, et ce davantage pour la disparité croisée qu'homonyme (Hinkle & Connor, 2001). La préférence pour la disparité croisée viendrait du fait que la voie ventrale mettrait l'emphase au premier plan des objets ou aux parties des objets les plus près de l'observateur (Hinkle & Connor, 2001).

En bref, bien qu'il y ait une dissociation entre les voies ventrale et dorsale, il semble que la voie ventrale utilise les informations de profondeur binoculaire et monoculaires pour extraire l'information sur les formes, et que la voie dorsale soit aussi impliquée dans la représentation des formes.

Objectifs

Comme le montrent les paragraphes précédents, plusieurs études se sont intéressées au rôle de diverses informations de profondeur, et entre autre à la stéréoscopie, dans la représentation de formes. Les résultats sont mitigés et il n'y a à ce jour pas de consensus sur le sujet. Il faut souligner également qu'au meilleur de notre connaissance, aucune étude n'a évalué la contribution de l'information de profondeur stéréoscopique dans la reconnaissance de forme et l'invariance au point de vue en utilisant la stéréoscopie inversée. Cela permettrait pourtant d'évaluer davantage l'implication de la stéréoscopie dans la représentation des formes. En effet, l'information donnée par la stéréoscopie inversée entrerait en conflit avec l'information amenée par les indices monoculaires. Si l'information binoculaire est utilisée par les processus impliqués dans la reconnaissance et la constance de forme, la stéréo inversée devrait nuire aux performances par rapport à une présentation stéréoscopique et même à l'absence de stéréoscopie. Aussi, aucune étude n'a évalué l'intégration de l'ombrage et de la stéréoscopie dans la perception du relief en évaluant l'impact de la disparité binoculaire sur l'effet de direction d'illumination.

Le but du premier article est d'évaluer l'impact de l'information de profondeur stéréoscopique dans la reconnaissance de forme lorsque l'information de profondeur monoculaire disponible est riche (i.e. ombrage, contours, perspective linéaire et occlusion), la question étant : Y a-t-il place pour une contribution de la stéréoscopie à la perception de la forme dans un tel contexte? Un second but était de vérifier si la stéréoscopie inversée ajoute effectivement de la sensibilité pour l'évaluation de l'impact de l'information binoculaire dans la perception des formes. L'expérience a utilisé une tâche de reconnaissance de formes (i.e. déterminer si un stimulus est familier ou nouveau suite à une phase d'apprentissage) avec une manipulation inter-sujets du mode de présentation. Les stimuli pouvaient être présentés avec disparité binoculaire, avec disparité inversée ou la même vue pouvait être présentée aux deux yeux. Ce dernier mode de présentation a été utilisé plutôt qu'une présentation monoculaire des stimuli, car elle peut être comparable aux conditions de disparité et de disparité inversée puisque l'information utilisée par les processus impliqués dans la tâche provient des deux yeux. Cependant, le fait que la même vue soit présentée aux deux yeux ne correspond pas à une absence de disparité entre les images rétiniennes des deux yeux. Cette condition peut donc amener un certain conflit entre l'information binoculaire et les informations monoculaires.

Cependant, l'ampleur du conflit est moins grande qu'avec la disparité inversée et comparer ces différentes conditions permet de voir l'impact de l'incongruence entre les indices de profondeur. Les stimuli étaient des formes 3D constituées de deux parties volumétriques. D'après les théories de Biederman (1987) et de Pizlo (2008), l'information provenant des contours est suffisante à elle seule pour reconnaître ces formes. Toutefois, si la stéréoscopie est utilisée par les processus sous-tendant la reconnaissance de forme, on peut s'attendre à ce que les performances avec stéréoscopie normale (3D) soient meilleures qu'avec stéréoscopie inversée (3D inversée) ou absence de disparité entre les vues présentées aux deux yeux (2D). De plus, si la stéréoscopie est utile à la reconnaissance de forme, on peut s'attendre à ce que la disparité inversée nuise aux performances par rapport à la 2D. Cependant, il est aussi possible que l'inconsistance entre les informations binoculaire et monoculaires fasse en sorte que la stéréoscopie inversée soit ignorée. Dans un tel cas, on pourrait s'attendre à ce que les performances soient meilleures pour la 3D que pour la 2D et la 3D inversée, sans différence entre ces dernières.

Le but du deuxième article est de vérifier comment l'ombrage et la stéréoscopie sont intégrés dans la perception du relief. L'expérience a utilisé une tâche jugement du relief (i.e. déterminer si des formes présentées sont convexes ou concaves). Les formes pouvaient être présentées sans disparité entre les vues présentées à chaque œil (pour les mêmes raisons que dans l'article 1), avec disparité binoculaire normale ou avec disparité binoculaire inversée, et l'illumination pouvait provenir du haut à gauche, du haut à droite, du bas à gauche ou du bas à droite. Sur la base des études d'Adams (2007), Connor (2001), Gerardin, Montalembert et Mamassian (2007), Kleffner et Ramachandran (1992) et Ramachandran (1988), un biais pour la perception de formes convexes est attendu lorsque l'éclairage vient d'en haut. De plus, étant donné les observations de Mamassian et Goutcher (2001) et de Sun et Perona (1998), on peut s'attendre à ce qu'elles soient perçues convexes plus souvent lorsque l'éclairage vient du haut à gauche qu'à droite. L'hypothèse d'une indépendance de la contribution de l'ombrage et de la disparité binoculaire prédit que les effets du mode de présentation et de direction d'illumination seront additifs. Cependant, si le traitement de ces informations n'est pas indépendant, une des deux informations devrait dominer l'autre, ou à tout le moins moduler les effets de l'autre, tel que le suggèrent les résultats de Braunstein (1986), Mamassian et Landy (2001), et Stevens et Brookes (1998). Par exemple, on pourrait constater l'absence d'impact de la direction d'illumination lorsque les formes sont présentées avec stéréoscopie, quelle soit normale ou inversée.

Les objectifs du troisième article sont d'évaluer l'impact de l'information de profondeur stéréoscopique dans l'invariance à l'orientation de formes non-familières qui ne répondent pas aux conditions d'invariance à l'orientation de Biederman et Gerharstein (1993), et d'évaluer l'impact du développement de l'expertise perceptuelle sur la contribution de la stéréoscopie. L'expérience consiste en une tâche d'appariement séquentiel de stimuli en forme de trombones tordus ayant subi des rotations 3D qui ont été présentés sans disparité entre les vues présentées aux deux yeux (pour les mêmes raisons que dans les articles 1 et 2), avec disparité binoculaire ou avec disparité binoculaire inversée. L'expérience a été menée à trois reprises se distinguant par certains aspects de la procédure appliquée. La première fois, le facteur «mode de présentation » (i.e. sans disparité entre les vues (2D), 3D ou 3D inversée) était intra-sujet, la deuxième fois intra-sujet avec un changement de stimuli au milieu de l'expérience, et la troisième fois inter-sujet. Ces manipulations ont permis de vérifier si la familiarité avec les stimuli influence la manière d'utiliser l'information stéréoscopique, et si le contexte d’exposition à différents modes de présentation affecte leurs effets. Étant donné les résultats de Bennett et Vuong (2006), Burke, (2005), et Burke et al. (2007) décrits précédemment, on pourrait s’attendre à ce que les coûts de rotation soient plus faibles si les formes sont présentées avec stéréoscopie normale que sans disparité entre les vues ou avec stéréoscopie inversée. Les mêmes résultats sont prédits par la théorie de reconnaissance par composantes de Biederman (Biederman, 1987; Biederman & Gerhardstein) puisque les stimuli en forme de trombones tordus ne respectent pas les conditions d’invariance à l’orientation. Par contre, d’après la théorie de Pizlo (2008) on peut s’attendre à ce que la stéréoscopie ne contribue pas à réduire les coûts de rotation. Qui plus est, si, comme il a été montré par Ichikawa et Egusa (1993), lorsque la stéréoscopie et les informations monoculaires sont incongruentes, plus de poids est donné aux informations monoculaires, on peut s’attendre à ce qu’il n’y ait pas de différence dans les coûts de rotation entre les modes de présentations 2D (absence de disparité entre les vues) et 3D inversée. En effet, dans toutes les conditions, les mêmes informations monoculaires étaient présentes, et si plus de poids est donné aux informations monoculaires qu’à la stéréoscopie lorsqu’il y a incongruence entre les différents types d’information, on pourrait s’attendre à ce que les coûts de rotation ne diffèrent pas entre ces deux dernières conditions d’exposition.

Article 1

STEREOPSIS CONTRIBUTES TO VISUAL SHAPE RECOGNITION

Mercédès Aubin & Martin Arguin

Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal

Contribution des coauteurs :

M.Aubin., & M. Arguin. : Conception du projet

M. Aubin. : Recrutement des participants

M. Aubin. : Collecte des données

M. Aubin. : Analyse des résultats

M. Aubin. : Rédaction de l'article

M. Aubin. : Interprétation des résultats

M. Aubin, & M. Arguin : Révision du manuscrit

M. Arguin : Financement

Abstract

We examined the role of stereopsis in visual shape perception using a recognition task. The stimuli were presented with no disparity between the views presented at each eye (2D), normal binocular disparity (3D) or reversed disparity (reversed 3D). Recognition performance was better with 2D and 3D displays than with reversed 3D. The absence of difference between the 2D and 3D conditions is probably due to a floor effect, and the difference between these two conditions and the reversed 3D one indicates that stereopsis contributes to shape recognition.

Understanding the representation of three-dimensional (3D) shape is fundamental to theories of vision (e.g. Biederman, 1987; Leek, Reppa, Rodriguez & Arguin, 2009). One important outstanding issue concerns the contribution of stereoscopic depth to shape perception.

It has long been known that it is possible to perceive 3D shapes using only information carried by binocular disparity. Indeed, Julesz (1960, 1971) showed that it is possible to recognise 3D shapes presented as random-dot stereograms, which are devoid of monocular shape cues and thus defined exclusively by binocular disparity. Some later studies also demonstrated that it is possible to recognise 3D curved surfaces defined by random-dot stereogram (De Vries, Kappers & Koenderink, 1993; Uttal, Davis & Welke, 1994; Uttal, Davis, Welke & Kakarala, 1988; Vreven, 2006.)

On the other hand, the perception of 3D shape does not require binocular disparity. Indeed, Ramachandran (1988) showed that it is possible to perceive 3D shapes only defined by shading. Furthermore, line drawn objects and photographs of objects can be recognised easily (Biederman & Ju, 1988). It seems to be even possible to achieve shape constancy with shapes defined only by their edges (Biederman, 1987; Biederman & Gerhardstein, 1993; Hummel & Biederman, 1992). While 3D shape perception can be achieved on the basis of monocular or binocular information only, a debate remains as to whether stereopsis is useful for shape perception when rich monocular information is available to define the shapes.

According to current structural theories (Biederman, 1987; Biederman & Bar, 1999; Biederman & Gerhardstein, 1993; Hummel & Biederman, 1992; Marr, 1982; Marr & Nishihara, 1978; Peissig, Wasserman, Young, & Biederman, 2002; Pentland, 1985), internal shape representations encode the 3D structure of objects in terms of a collection of volumetric primitives and their connectedness. Biederman proposed that non-accidental properties, which are 3D shape properties shared across viewpoints, would be the basis for object recognition. As long as the non-accidental properties remain visible, shape representation should be orientation invariant. This theory thus suggests that the implicit 3D cues provided by edges are sufficient to create shape constancy.

In the same vein, Pizlo and colleagues (Chan, Stevenson, Li & Pizlo, 2006; Pizlo, 2008; Pizlo, Li & Francis, 2005; Pizlo, Li, Steinman, 2008; Pizlo & Stevenson, 1999) argue that shape representation and shape constancy are based on a priori monocular simplicity constraints. Consequently, binocular depth cues are neither necessary nor sufficient for shape constancy. In support of this view, Pizlo et al. (2005) have reported that when monocular simplicity constraints and stereoscopic depth cues are inconsistent, binocular cues are ignored. Furthermore, the performance of models of monocular shape representation in a sequential discrimination task correlate with human monocular and binocular performances whereas the binocular models performances do not (Chan & al, 2006). A more recent study by Li and Pizlo (2011) supports the fact that edges and stereopsis are useful to achieve shape constancy, but that when the performance is at chance with monocular cues only, they remain at chance when binocular information is added.

However, some studies have shown that for stimuli like faces, bent paperclips and other tubular objects, stereopsis contributes significantly to shape constancy (Bennett & Vuong, 2006; Burke, 2005; Burke, Taubert & Higman, 2007). As far as objects are concerned, the classes of stimuli that have been sampled thus far are rather limited and they fall outside the range of shapes that are relevant for either Biederman's or Pizlo's theories: i.e. they cannot be defined by geon structural descriptions (cf. Biederman) and do not respond to simplicity constraints (cf. Pizlo). Thus, it is possible that the findings currently available do not generalize to other classes that are visually closer to what we encounter in daily life, and so the role of stereopsis in shape perception remains unclear when the shapes contain rich monocular information.

To explore the contribution of stereo to shape perception, reversed stereo appears as a relevant exposure condition that has been poorly explored so far. Wheatsone (1852) used the pseudoscope (a stereoscope that presents the image normally going to the left eye to the right eye, and vice-versa) and observed that the objects seen through it had reversed depth. The more recent studies that have examined the effect of reversed stereopsis on visual perception (Ichikawa & Egusa, 1993; Ichikawa, Egusa, Nakatsuka, Amano, Ueda & Tashiro, 2003; Shimojo & Nakajame, 1981; Wheatsone, 1852; Yellot & Kaiwi, 1979) have focused on the adaptation effect resulting from the long term wearing of right-left reversing spectacles which reversed binocular disparities. This adaptation effect causes depth inversion once the spectacles are removed, and it also alters the weight of the different depth cues (Ichikawa & Egusa, 1993). Indeed, participants ended up ignoring binocular disparity altogether and using occlusion and linear perspective to a greater degree to make depth judgements than before the adaptation.

However, few studies have yet been conducted to determine the impact of reversed stereopsis on shape recognition. Such a study is relevant however, because if stereopsis contributes to shape recognition, reversed disparity should alter performance compared to normal disparity or the absence of disparity since the information carried by stereopsis would be inconsistent with others cues like shading, linear perspective, edges and occlusion.

It is worth noting that some studies have evaluated the impact of the inconsistency of different cues to shape. Stevens, Lees and Brooke (1991) showed that when information from stereopsis and edges is inconsistent, participants tended to rely on one cue and to ignore the other for surface perception. Norman and Todd (1995) found that for surface perception, when depth from motion and depth from stereopsis were inconsistent, one of the two cues was used and the other suppressed, depending on which was the most informative. However, there were important between-subject variations. For surface slant evaluation, texture seems to be the preferred source of information, when there is inconsistency with stereopsis (Saunder & Backus, 2006). When stereopsis and monocular cues are inconsistent (occlusion and texture gradient), the monocular cues override stereopsis in direction of rotation judgements and surface slant evaluation (Braunstein, Andersen, Rouse & Tittle, 1986; Stevens & Brookes, 1988.) Nevertheless, the impact of the inconsistency of the different cues to shape has never been studied specifically in the context of 3D shape recognition.

In the present study, we intend to determine the role of stereopsis in visual 3D shape recognition, when rich monocular information is available (i.e. shading, edges, linear perspective, and occlusion). The experiment used a shape recognition task (i.e. determining if a stimulus is new or familiar, following a learning phase) with a between-subjects manipulation of display mode (normal binocular disparity (3D), no disparity between the views presented to the eyes (2D) and reversed disparity (reversed 3D)). The stimuli were 3D shapes that contained two putative volumetric parts. According to Biederman (1987) and Pizlo (2008), edges constitute the only necessary source of information to achieve shape representation and thus, display mode should not affect performance. However, if stereopsis is involved in the processes leading to shape recognition, we should expect better performance with normal stereopsis than with the other display modes. Also, if stereopsis contributes to 3D shape recognition, we should expect reversed stereopsis to decrease performance compared to the 2D display mode, since the discrepancy between monocular and binocular information is more important in the reversed 3D condition than in the 2D condition. However, if the inconsistency between stereopsis and monocular leads participants to ignore stereopsis altogether, performance should be better in participants exposed to normal stereopsis than with either of the other two display modes, which should not differ from one another.

Method

Participants

Three groups of eight participants (15 females and 9 males, 23 right-handers and 1 left-hander) aged between 19 and 34 years old took part in the study. All had normal or corrected vision.

Stimuli and material

The stimuli were presented over a white background on a 16-inch Compaq monitor of 1024 x 768 pixels resolution. Participants viewed stimuli through a mirror stereoscope by Stereo Aids, which presented the right half of the screen to the right eye, and the left half of the screen to the left eye. The screen was split in half: the stimuli for the right eye were centered in the right half, and those for the left eye centered in the left half. The experiment was run on a Pentium 4 computer, and its progress and registration of the observer's performance were controlled by the E-Prime software. Participants responded by pressing the buttons of a mouse.

The stimuli were created using the 3D Studio MAX® program from Autodesk Media & Entertainment (CA, USA.) They correspond to twelve unfamiliar 3D shapes that contained two putative volumetric parts each, replicating the stimuli used by Leek et al. (2009). They were viewed dichoptically through a mirror stereoscope. They had a maximal spatial extent of 13.8 x 13.8 degrees of visual angle from the viewing distance of 60 cm. A rotation of 5.42 degrees around the vertical axis was applied to the stimuli to create distinct views for the left and right eyes for the stereo and reversed stereo conditions (simulating an interocular distance of 56 mm)[1]. For the 2D condition, the same image was presented to both eyes.

Insert Fig. 1 near here

Design

The experiment proceeded in three phases: study, practice and test. In the study phase the participants memorized six objects. Half of the participants from each group memorized the stimuli from subset A (Fig 1), and the other memorized subset B. The non-studied stimuli served as the "novel objects" for the test phase. During the study phase, participants were shown each object individually for an unlimited duration. The next object was presented when the participant indicated he had memorized the shape. During the practice phase, participants completed blocks of 18 trials in which the six learned objects and three new ones (not used

---

[1] Note that the optimal way to create stereo pair would be to use two virtual viewpoints, with one horizontally displaced relative to the other in a direction parallel to the display plane. Renderings using this method were compared to those used in the experiment. The correlations for the left and right-eye views were both greater than .99, indicating that the images were almost perfectly identical across the two rendering methods.

later in the test phase) occurred twice each. Participants indicated whether the object was familiar or new. The test phase started when they obtained at least 15 correct responses in a block (i.e. 83% correct). During the test phase, participants had to determine if the object displayed on each trial was familiar or new.

Procedure

Participants first completed the study and the practice phases. In the test phase, trials began with a fixation cross displayed in the middle of the screen for 750 ms, followed by a delay of 250 ms and then by the target. The target remained visible until the participant's response. Incorrect responses were immediately followed by an error message ("incorrect") displayed on the screen for 1000 ms. There was a delay of 500 ms prior to the next trial. Participants were asked to respond as fast and accurately as possible. Each familiar and novel object was presented 10 times. Thus, there were 60 "familiar" trials and 60 "new" trials for a total of 120 trials. Trial order was randomized within a unique test block.

Results

Response times (RTs) that were more than 2.5 standard deviations away from a participant's average for a given condition were eliminated, resulting in the exclusion of a total of 77 data points for the complete experiment (2.6% of trials). The error trials were also excluded from the RTs analysis (5.8% of trials).

RTs and error rates (ERs) analysis

Two-way mixed ANOVAs including the factors of response (familiar or new) and display mode (2D, 3D, reversed 3D) were carried out on the correct RTs and the ERs. The RTs analysis revealed no significant effect (all *p*'s > .67). The ERs analysis showed no significant effect of response (*F* (1, 21) < 1), and no significant response x display mode interaction (*F* (2, 21) < 1). However there was a significant display mode effect (*F* (2, 21) = 7.58, *p* <.01, η^2 = .42; see Fig. 2). Indeed, the ERs were lower for the 3D (*F* (1, 14) = 13.34, *p* <.005, η^2 = .51) and 2D (*F* (1, 14) = 7.36, *p* = .017, η^2 = .35) conditions than for the reversed 3D condition (alpha of .017 after Bonferroni correction for multiple comparisons). However, there was no significant difference between the 2D and the 3D conditions (*F* (1, 14) < 1).

Discussion

In brief, the results show that there was no performance difference in the recognition of shapes presented with normal binocular disparity, or with no disparity between the views presented at each eye. However reversed stereopsis had an incremental effect on error rates.

The performance cost observed with reversed stereopsis relative to normal stereopsis or 2D displays probably indicates that stereo is a depth cue that contributes to visual shape recognition. We attribute the lack of difference between the 2D and 3D conditions to a floor effect. Indeed, the monocular depth information available was rather rich, thereby giving little room for additional gains from stereopsis. Furthermore, the ERs could barely be lower in the 2D and 3D conditions as the participants made an average of 4.2 errors per 120 trials.

It is worth underlining that the present evidence in support of a contribution of stereopsis to visual shape perception rests on our use of reversed-3D displays. Had we not used this condition, our conclusions would have been opposite and incorrect. This indicates the relevance of using reversed stereo in assessments of the contribution of stereoscopic information to a particular visual function.

The fact that reversed stereopsis brought a disadvantage indicates that for the recognition of the class of stimuli used in this experiment, when binocular and monocular information are inconsistent, binocular information is taken into account and not ignored or suppressed as would be expected from the results of Braunstein et al. (1986), Norman and Todd (1995), Saunders and Backus (2006), and Stevens and Brookes (1988). It appears that in the present experiment participants tried to integrate the inconsistent binocular information with the monocular depth cues, even if it was at the expense of their performance. It should be noted however, that the monocular depth cues studied in Braunstein & al., Norman & Todd (1995), Saunders & Backus (2006), and Stevens & Brookes are occlusion and velocity, motion or texture gradients, in contrast to shading and occlusion in the present study. This suggests that different monocular depth cues may differ in the degree to which they contribute to the representation of shape and in how they interact with stereo. A mitigating factor that may also be important in determining the contribution of different depth cues is the type of stimuli used. Thus, the stimuli used by Braunstein et al. and in Stevens and Brookes are line drawings, those used by Norman and Todd were points of light, and those of Saunders and Backus were planar surfaces defined by texture gradient, in contrast to the shaded complex objects used here. Another important factor concerns the constraints of the task and the nature of the stimulus interpretation required. Indeed, the current task required participants to recognise complex shapes, instead of evaluating the rotation direction of a sphere (Braunstein & al.), adjusting the depth of surfaces (Norman & Todd), evaluating surfaces slant (Saunders & Backus), or judging whether a probe was perpendicular to a surface (Stevens & Brookes). We believe that the markedly different constraints of the present task and the nature of the stimulus interpretation required is probably what led to the difference in how depth information was used.

It was already well known that stereopsis contributes to shape perception when binocular disparity is the only available information (De Vries, Kappers & Koenderink, 1993; Julesz, 1960; 1971; Uttal, Davis & Welke, 1994; Uttal, Davis, Welke & Kakarala, 1988; Vreven, 2006.). However, the contribution of stereopsis when monocular depth cues are readily available was less clear. Biederman's (1987) and Pizlo's (2008) theories claim that the information provided by edges is sufficient on its own to achieve an efficient representation of visual shapes, and that stereopsis is not used. This claim however, is applicable only to shapes that are relevant under the said theories: i.e. they can be defined by geon structural descriptions (cf. Biederman) and they respond to simplicity constraints (cf. Pizlo). The shapes used here obey these constraints and nevertheless demonstrate sensitivity to stereoscopic depth information. The present findings thus falsify the statements made previously by Biederman and Pizlo in this regard. They also provide support for the findings of Bennett and Vuong (2006), Burke (2005), and Burke et al. (2007) which showed that stereopsis facilitates visual recognition by contributing to shape constancy. The stimuli used in these studies did not meet the constraints imposed by Biederman's and Pizlo's theories but they offered rich monocular depth cues. Indeed, those used by Burke (2005) and Burke et al. (2007) were photographs, and the ones used by Bennett and Vuong, contained shading and occlusion. Our stimuli also presented rich monocular information and stereopsis still contributed to their recognition. However, it is worth reminding that without the reversed stereopsis condition, this contribution would have been missed.

Limits

We took for account that the absence of difference between the 2D and 3D conditions was due to a floor effect and that stereo is implicated in shape recognition. However, one could also propose another interpretation of our result. Indeed, it is possible that 2D and 3D lead to identical performance because stereo does not bring any advantage or is not used to achieve the recognition of the kind of shape used here, but that it is taken into account only when it is very inconsistent with monocular information. Our results don't allow us to settle the question, and to determine which hypothesis is most likely to be true. To examine the question, it would be interesting to remake the experiment but with a higher difficulty level, by reducing the contrast of the shape to make them harder to see for example, and to verify if the performance are still equal for the 2D and 3D conditions or if the performances are better in the 3D condition than in the 2D one.

Further studies

Finally, it is well known that the ventral pathway processes visual objects, forms, texture and color, and that the dorsal pathway is implicated in spatial localisation and in the visual guidance of action (Goodale & Milner, 1992; Goodale et Westwood 2004; Ungerleider & Mishkin, 1982). However it has been found that some neurons in the ventral pathway of the macaque show sensitivity to binocular disparity (Hinkle & Connor, 2001; Uka, Tanaka, Yoshiyama, Kato & Fujita, 2000), and that in humans, V5, which is part of the dorsal pathway, is sensitive to 3D shapes created by the combination of different depth cues (Welchman, Deubelius, Conrad, Bülthoff and Kourtzi, 2005). These results are consistent with the contribution of stereopsis to shape recognition that we measured since both pathways seem to integrate the depth cues to achieve shape representation. The two pathways possibly work together to combine the different depth cues and to achieve shape recognition, and their conjoint contribution has hardly been explored.

Conclusions

To conclude, the present experiment showed that stereopsis contributes to the processes underlining shape recognition, even when the monocular depth information is rather rich. It also demonstrated that reversed stereopsis is a relevant condition to further explore the role of binocular disparity in visual perception.

Acknowledgements

The research reported here was supported by grant no. 203613 from the Natural Sciences and Engineering Research Council of Canada to Martin Arguin.

References

Bennett, D.J., & Vuong, Q.C. (2006) A stereo advantage in generalizing over changes in viewpoint on object recognition tasks, *Perception and Psychophysics, 68,* 1082-1093.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding, *Psychological Review, 94,* 115-147.

Biederman, I. & Bar, M. (1999) One-shot viewpoint invariance in matching novel objects, *Vision Research, 39,* 2885-2899.

Biederman I, & Gerharstein, P.C. (1993) Recognizing depth-rotated objects: Evidence and conditions for three-dimensional view-point invariance, *Journal of experimental Psychology: Human perception and performance, 19*, 1162-1182.

Biederman, I., & Ju, G. (1988) Surface versus edge-base determinants of visual recognition, *Cognitive Psychology, 20,* 38-64.

Braunstein, M.L., Andersen, G.J., Rouse, M.W., & Tittle, J.S. (1986) Recovering viewer-centered depth from disparity, occlusion and velocity gradients, *Perception & Psychophysics, 40*, 216-224.

Burke, D. (2005) Combining disparate views of objects: Viewpoint costs are reduced by stereopsis, *Visual Cognition, 12,* 201-213.

Burke, D., Taubert, J. & Higman, T. (2007) Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces, *Vision Research, 47*, 2164-2169.

Chan, M.W., Stevenson, A.K., Li, Y., & Pizlo, Z. (2006) Binocular shape constancy from novel views: The role of a priori constraints, *Perception & Psychophysics, 68,* 1124-1139.

De Vries, S.C., Kappers, A.M.L., & Koenderink, J.J. (1993) Shape from stereo: A systematic approach using quadratic surfaces, *Perception & Psychophysics, 53,* 71-80.

Goodale, M.A. & Milner, A.D. (1992) Separate visual pathways for perception and action, *Trends in Neuroscience*, *15*, 20-25.

Goodale, M.A. & Westwood, D.A.(2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action, Current Opinion in *Neurobiology, 14*, 203-211.

Hinkle, D.A. & Connor, C.E. (2001) Disparity tuning in macaque area V4, *Neuroreport, 12*, 365-369.

Hummel, J.E. & Biederman, I. (1992) Dynamic Binding in a Neural Network for Shape Recognition, *Psychological Review, 99,* 480-517.

Ichikawa, M. & Egusa, H. (1993) How is depth perception affected by long-term wearing of left-right reversing spactacles?, *Perception, 22*, 971-984.

Ichikawa, M., Eguase, H., Nakatsuka, M., Amano, J., Ueda, T., Tashiro, T. (2003), Modification of depth and distance perception caused by long-term wearing of left-right reversing spectacles, *Perception, 32*, 131-153.

Julesz, B. (1960) Binocular depth perception of computer-generated patterns, *The Bell System Technical Journal, 39,* 1125-1162.

Julesz, B. (1971) Foundations of cyclopean perception, Chicago: University of Chicago Press.

Leek, E.C., Reppa, I., Roderiguez, E. & Arguin, M. (2009) Surface but not volumetric part structure mediates three-dimensional shape representation: Evidence from part-whole priming, *The Quaterly Journal of Experimental Psychology, 62*, 814-830.

Li, Y., & Pizlo, Z. (2011) Depth cues versus the simplicity principle in 3D shape perception, *Topics in Cognitive Science, 3,* 667-685.

Marr, D. (1982). Vision. San Francisco: Freeman.

Marr, D., & Nishihara, H.K. (1978) Representation and recognition of three dimensional shapes, *Proceeding of the Royal Society of London, B, 200*, 269-291.

Norman, J.F. & Todd, T.T. (1995) The perception of 3-D structure from contradictory optical patterns, *Perception & Psychophysics, 57,* 826-834.

Pizlo, Z. (2008) 3D Shape: Its Unique Place in Visual Perception, MIT Press, Cambridge.

Pizlo, Z., Li, Y., & Francis, G. (2005) A new look at binoculair stereopsis, *Vision Research, 45*, 2244-2255.

Pizlo, Z., Li, Y., & Steinman, R.M. (2008) Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose, *Spatial Vision, 21,* 495-508.

Pizlo, Z. & Stevenson, A. (1999) Shape constancy from novel views, *Perception & Psychophysics, 61,* 1299-1307.

Peissig, J. J., Wasserman, E. A., Young, M. E., & Biederman, I. (2002). Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons. *Vision Research, 42*, 2051–2062.

Pentland, A. (1985). On describing complex surface shapes. *Image and Vision Computing, 3*, 153–162.

Ramachandran, V.S. (1988) Perception of shape from shading, *Nature, 331*, 163-166.

Saunders, J.A. & Backus, B.T. (2006) Perception of surface slant from oriented textures, *Journal of Vision, 6*, 882-897.

Shimojo, S., & Nakajama, Y. (1981) Adaptation to the reversal of binocular depth cues : Effects of wearing left-right reversing spectacles on stereoscopic depth perception, *Perception, 10*, 391-402.

Stevens, K.A., & Brookes, A., (1988) Integrating stereopsis with monocular interpretations of planar surfaces, *Vision Research, 28*, 371-386.

Stevens, K.A., Lees, M., et Brookes, A. (1991) Combining binocular and monocular features, *Perception, 20*, 425-440.

Uka, T., Tanaka, H., Yoshiyama, K., Kato, M. & Fujita, I. (2000) Disparity selectivity of neurons in monkey inferior temporal cortex, *Journal of Neurophysiology, 84*, 120-132.

Ungerleider, L.G., & Mishkin, M. (1982) Two cortical systems, *Analysis of Visual Behavior,* Cambridge, MA, MIT Press, 549-586.

Uttal, W.R., Davis, N.D., & Wlke, C. (1994) Stereoscopic perception with brief exposures, *Perception & Psychophysic, 56*, 599-604.

Uttal, W.R., Davis, N.S., Wleke, C., & Kakarala, R. (1988) The reconstruction of static visual forms from sparse dotted samples, *Perception & Psychophysics, 43*, 223- 240.

Vreven, D. (2006) 3D shape discrimination using relative disparity derivatives, *Vision Research, 46*, 4181-4192.

Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H. & Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortex, *Nature Neuroscience, 8,* 820-827.

Wheatsone, C. (1852), Contributions to the physiology of vision part II. On some remarkable, and hitherto unobserved, phenomena of binocular vision, *Abstract of the papers communicated to the Royal Society of London, 6*.

Yellott, J.I., & Kaiwi, J.L. (1979) Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective, *Perception, 8*, 135-142.

Figure captions

FIG. 1: The twelve stimuli used in the experiment. A) One set of stimuli, which were learned by half the participants and which served as novel items for the other half. B) The other set of stimuli which served a complementary role to set A.

FIG. 2: The ERs as a function of display condition. The error bars represent the standard deviations.

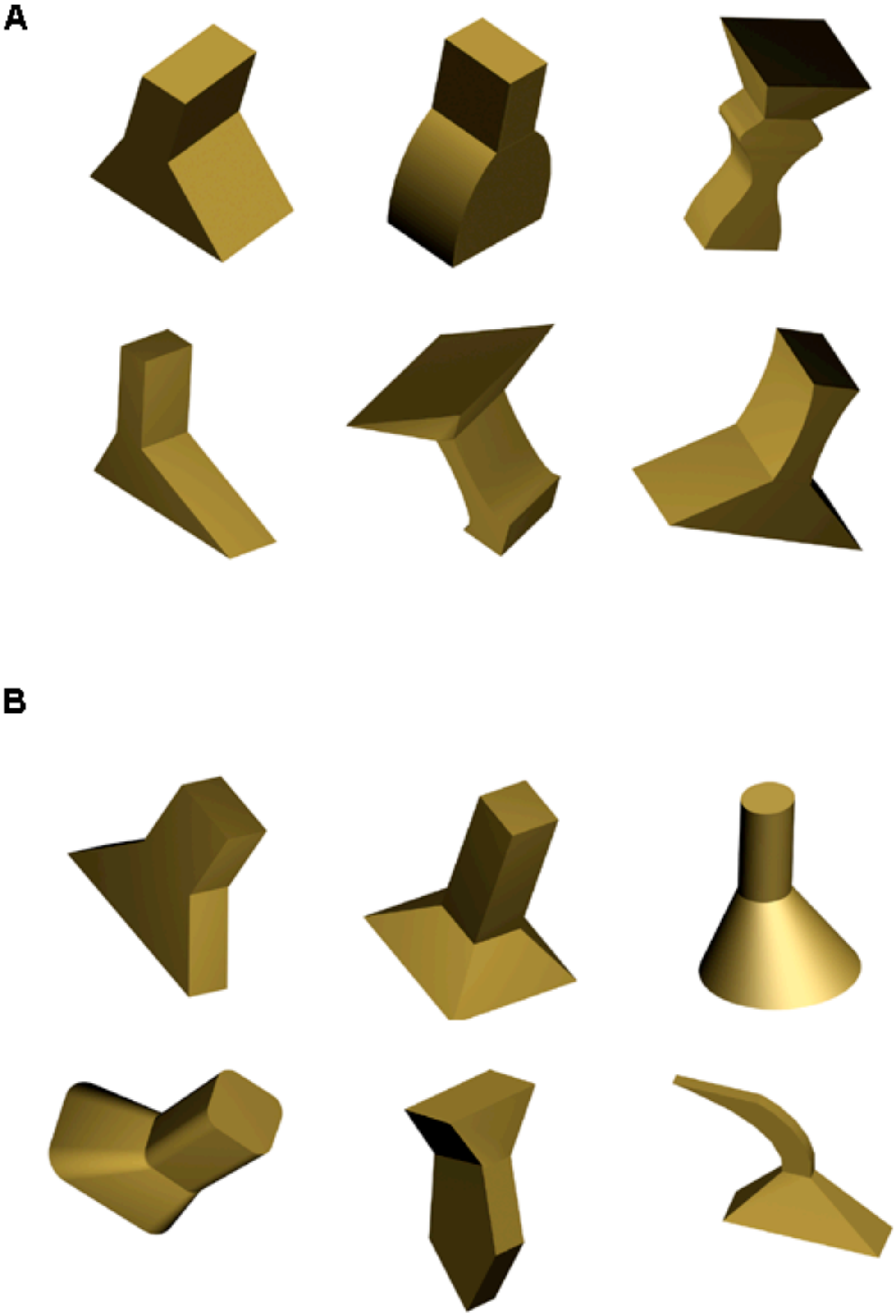


Fig. 1

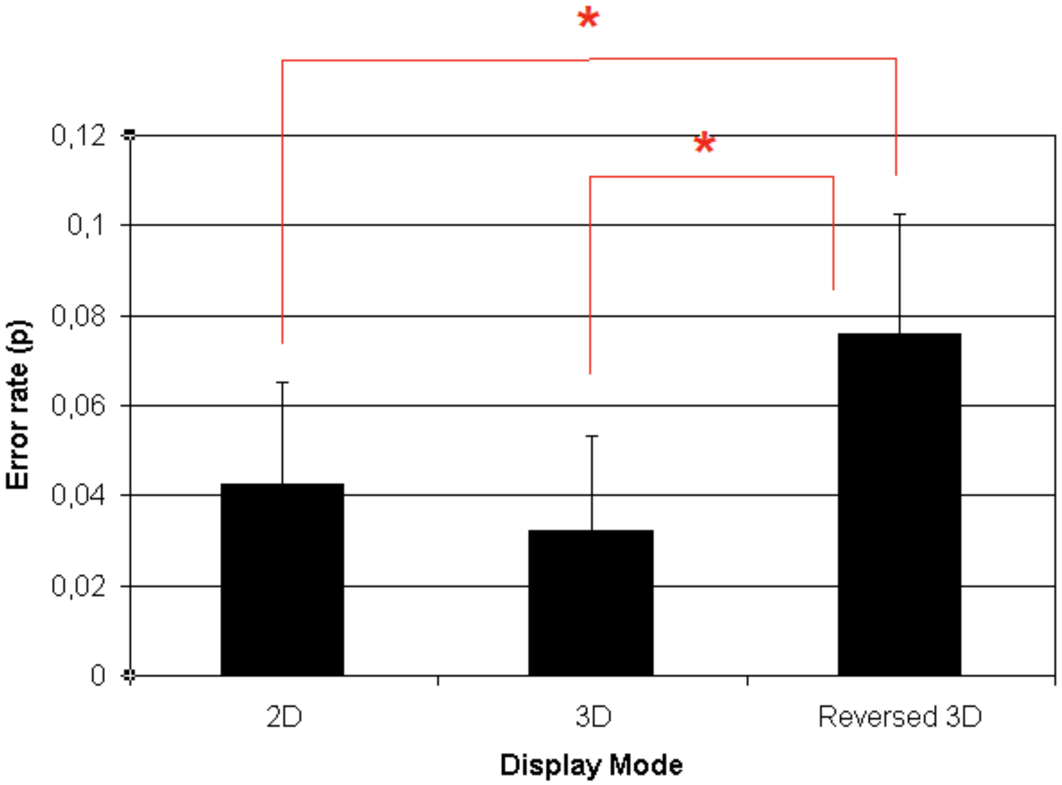

*
*
0,12
0,1
0,08
0,06
0,04
0,02
0
Error rate (p)
2D
3D
Reversed 3D
Display Mode

Fig.2

Article 2

STEREO AND SHADING CONTRIBUTE INDEPENDENTLY TO SHAPE CONVEXITY/CONCAVITY DISCRIMINATION

Mercédès Aubin & Martin Arguin

Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal

Article accepté par Perception

Contribution des coauteurs :
M.Aubin. : Conception du projet
M. Aubin. : Recrutement des participants
M. Aubin. : Collecte des données
M. Aubin. : Analyse des résultats
M. Aubin. : Rédaction de l’article
M. Aubin. : Interprétation des résultats
M. Aubin, & M. Arguin : Révision du manuscrit
M. Arguin : Financement

Abstract

The present study examined the joint contribution of shading and stereopsis to the perception of shape convexity/concavity. The stimuli were the images of a synthetic convex 3D shape seen from viewpoints leading to ambiguity as to its convexity. Illumination either came from above, or below and from the right or the left, and stimuli were presented dichoptically with either normal binocular disparity, reversed disparity, or no disparity. Participants responded “convex” more often when the lighting came from above than from below. Also, participants responded that the shape was convex more often with normal than with zero disparity, and more often with zero disparity than with reversed stereopsis. The effects of lighting direction and display mode were additive; i.e. they did not interact. This indicates that shading and stereopsis contribute independently to shape perception.

Understanding the representation of three-dimensional (3D) objects is fundamental to theories of vision. One unresolved issue concerns the integration of different depth cues in shape perception.

It is well known that we can perceive 3D shape from shading with the constraint that there is only one light source that illuminates the scene (Ramachandran, 1988). Shape from shading is also strongly constrained by a light-from-above prior; specifically, we assume that light comes from above which induces depth reversal when light comes from below (Adams, 2007; Brewster, 1826; Gerardin, de Montalembert & Mamassian, 2007; Kleffner & Ramachandran, 1992). This prior seems stronger with collimated lighting than with diffuse lighting (Langer & Bülthoff, 2000). It remains unclear whether there is also a favourite direction (left or right) with light coming from above. Sun and Perona (1998) and Mamassian and Goutcher (2001) reported a bias for light coming from above and left. Moreover, this bias correlates with handedness in the Sun and Perona study, but not in that of Mamassian and Goutcher. On the other hand, McManus, Buckman and Woolley (2004), who reported a preference for light from above in shape judgements, also found a leftward bias when stimuli were presented for 1 second or less, but not when they were presented for an unlimited duration (i.e. until response).

There is also a prior for convexity in shape from shading (Hill & Bruce, 1994; Langer & Bülthoff, 2001). Liu and Todd (2004) demonstrated that the convexity prior was stronger than the lighting direction biases with two tasks in which participants had to evaluate the sign and magnitude of surface curvature of shaded images. However, another study showed that the convexity prior is indeed stronger than the light-from-above prior in children, but not in adults, in whom the light-from-above prior seems to dominate (Thomas, Nardini & Mareschal, 2010). Furthermore, with visuo-haptic experience, it is possible to modify the convexity prior for both shape judgements and the visual search task (Champion & Adams, 2007). It is also possible to modify the light-from-above prior in shape evaluation tasks but not in the visual search task. From these observations, Champion and Adams (2007) argued that the convexity prior can be modified at a preattentive stage of processing (at which the pop-out effect occurs in visual search), but not the light-from-above prior.

The view-from-above prior also impacts shape perception (Mamassian & Landy, 1998; Reichel & Todd, 1990). Specifically, we tend to assume that the viewpoint from which we look at the object is from above. Mamassian and Landy (2001) have studied the interaction of light-from-above and view-from-above priors to explore the mechanisms subtending the integration of priors. By varying the contrast of the cues supporting each prior to modulate their reliability, they showed that when the “light-from-above” and view-from-above” priors suggest opposite interpretations, the conflict is resolved according to the reliability of the cues. They concluded that the more reliable cues lead to the attribution of a higher weight to their prior constraint (e.g. if shading is the most reliable cue, greater weight is given to the light-from-above prior). They note in this respect that priors act like depth cues.

There is indeed evidence indicating that depth cues are weighted depending on how reliable they are. Texture is more reliable to evaluate large than small slant. Stereopsis is also more reliable with large slant, but also with short viewing distances, as well as with the slant size effect modulated by viewing distance (Hillis, Watt, Landy & Banks, 2004; Knill & Saunders, 2003). These two cues seem to be optimally integrated and are given weights proportional to their reliability when it comes to slant discrimination and judgements (Hillis et al., 2004; Knill & Saunders, 2003). When these cues are in conflict, each receives different weights depending on which is the most informative for slant evaluation (Saunders & Backus, 2006). Norman and Todd (1995) found that for the perception of surface corrugation in depth, when stereo and motion contradict each other, the modality showing the more effective surface curvature direction (horizontal or vertical) was perceived and the other suppressed. Furthermore, some studies have shown that when stereopsis and monocular cues (occlusion and velocity or texture gradients) are inconsistent, the monocular cues override stereopsis (Braunstein et al., 1986; Stevens & Brookes, 1988). Bütlhoff and Mallot (1988; 1990) found that for local surface evaluation, if depth cues are in conflict, edge-based stereo overrides disparate shading and non-disparate shading. Furthermore, disparate shading inhibits non-disparate shading. They also found that when stimuli are lighted from below, stereo prevents depth reversal.

While a number of studies indicate that when different depth cues are available, some may override the others, the more common case is cue integration. Landy and colleagues have proposed the modified weak fusion (MWF) model as a general account of how depth cues are integrated (Landy, Maloney, Johnston, & Young, 1994). According to the MWF model, depth cues are weighted according to their reliability, availability, and consistency, and they are typically integrated linearly. The model can accommodate nonlinearities however, such as one cue vetoing another, in particular cases where cues are inconsistent or unreliable. In the same vein, Dosher, Sperling and Wurst (1986) demonstrated that in the perception of 3D structure, stereo and proximity luminance covariance (i.e. the increase in the edge intensity as a function of the proximity to the observer) are integrated linearly.

There is biological evidence that some neuronal substrates combine the different depth cues to achieve depth perception. Indeed, Tsutsui, Sakata, Naganuma and Taira (2001) show that in intraparietal sulcus of the macaque, some neuron populations respond selectively to surface orientation in depth defined by a texture gradient, regardless of the texture pattern, and most also respond selectively to the surface orientation in depth defined by random dots stereograms. Liu, Vogels and Orban (2004) found similar results with neurons of the infero-temporal cortex of the macaque that respond selectively to surface orientation in depth no matter whether the depth information was carried by texture or by disparity.

As noted previously, a number of studies have examined how different depth cues are integrated. However, no study has yet assessed the joint processing of stereopsis and direction of lighting in the perception of shape convexity/concavity. This is the purpose of the present study, wherein the impact of stereopsis will be studied not only by contrasting normal stereo displays to zero-disparity images, but also by including a reversed stereopsis condition. Reversed stereopsis is a potentially valuable test condition that has yet to be explored. Some studies have examined the adaptation effect to reversed stereopsis resulting from the long term wearing of right-left reversing spectacles, which thus reverse the sign of binocular disparities (Ichikawa & Egusa, 1993; Ichikawa, Egusa, Nakatsuka, Amano, Ueda & Tashiro, 2003; Shimojo & Nakajame, 1981; Yellot & Kaiwi, 1979). This adaptation led to a depth inversion after-effect once the spectacles were removed, and it also altered the weight of the different depth cues (Ichikawa & Egusa, 1993). Indeed, participants ended up ignoring binocular disparity altogether and using occlusion and linear perspective to a greater degree to make depth judgements than before the adaptation. Except for the studies with the hollow-mask illusion by Matthews, Hill and Palmisano (2011), the impact of reversed stereopsis on the perception of object relief without long term adaptation has yet to be investigated. A particular interest in using reversed stereopsis is that it maximizes the power of stereoscopic information manipulations to impact on performance. Indeed, if stereopsis, as shown by Bülthoff and Mallot (1988, 1990) helps prevent shape inversion with lighting from below, we should expect reversed stereopsis to amplify the likelihood of inversion with lighting from below and possibly to cause shape inversion even when shapes are lighted from above. Reversed stereopsis implies that the crossed disparities of a concave object viewed with normal stereopsis are transformed into uncrossed disparities. Crossed and uncrossed disparities are not equal and it has been proposed that they may be processed by distinct mechanisms (Mustillo, 1985). Since Bülthoff and Mallot (1988, 1990) only worked with crossed disparities, and that no study has yet examined the integration of uncrossed disparity with other depth cues, there is a possibility that reversed stereo may not produce effects symmetrical to those of normal stereo.

The goal of the present research was to determine if shading and stereopsis have independent (i.e. additive) or interactive (i.e. one factor modulates the impact of the other) contributions to shape perception. An experiment using a shape judgement task was used. Specifically, participants had to determine if the shape presented is convex or concave. The stimuli were the images of a convex 3D shape seen from viewpoints that lead to ambiguity as to the convexity of the shape (Fig. 1). Illumination either came from above, or below and from the right or the left, and stimuli could be presented with binocular disparity, reversed disparity or no disparity. This allowed us to determine whether shading and stereopsis are independent or interactive in their contribution to shape perception. For instance, based on the findings of Bülthoff and Mallot (1988, 1990), which suggest that stereopsis overrides shading, one would predict that when stereoscopic information (normal or reversed) is available, the impact of shading on the perception of convexity/concavity should be null or attenuated relative to zero-disparity displays. In contrast, if the effects of display mode and lighting direction do not interact, this would indicate that the two sources of information are treated independently for the determination of shape convexity/concavity.

Methods

Participants

Fourteen right-handed participants (4 males and 10 females) aged between 19 and 33 years old took part in the experiment. All were naive as to the purpose of the experiment, were neurologically intact, and had normal or corrected visual acuity, and a good stereo vision (assessed by the Stereo Fly Test). No particular selection was applied with respect to gender or level of education.

Material and Stimuli

The stimuli were presented over a white background on a 16-inch Compaq monitor of 1024 x 768 pixels resolution. The luminance of the screen was of 117 cd/m^2. Participants viewed stimuli through a mirror stereoscope by Stereo Aids which presented the right half of the screen to the right eye and the left half of the screen to the left eye. The screen was split in half: the stimuli for the right eye were centered in the right half, and those for the left eye centered in the left half. The experiment was run on a Pentium 4 computer, and its progress and registration of the observer's performance were controlled by the E-Prime software. Participants responded by pressing the buttons of a mouse.

The stimuli were created using the 3D Studio MAX® program from Autodesk Media & Entertainment (CA, USA) and rendered using orthographic projection. They all correspond to a unique lemon-like shape with flat extremities and a uniform gray surface which was presented from four different viewpoints (Fig. 1). The purpose of the flat extremity was to aid stereo matching; otherwise the stimulus information would have been too poor to lead to a strong 3D percept. The spatial extent of the stimuli was of 5.7 x 5.7 degrees of visual angle at the viewing distance of 60 cm. The shape could be lighted from above right, above left, below right and below left (22 degrees left or right of a vertical line running through the object's center, and 39 degrees above or below the horizontal, see Fig. 2.) Rotations of 5.42 degrees around the vertical axis were applied to the stimuli to create distinct views for the left and right eyes for the stereo and reversed stereo conditions, which simulates the effect of an inter-ocular distance of 5.6 cm for a 3D object viewed from 60 cm[2]. For the 2D display condition, stimuli with the same viewpoint were presented to both eyes. The Michelson contrast of the stimuli was of .99.

Insert Figs. 1 and 2 near here

Procedure

Participants indicated whether the target was convex or concave within a 3 (display mode: 3D, reversed 3D or 2D) x 2 (light from above or below) x 2 (light from the right or the left) repeated measures experimental design. At the beginning of each trial, a fixation cross

---

[2] Note that the optimal way to create stereo pair would be to use two virtual viewpoints, with one horizontally displaced relative to the other in a direction parallel to the display plane. Renderings using this method were compared to those used in the experiment. The correlations for the left and right-eye views were both greater than .99, indicating that the images were almost perfectly identical across the two rendering methods.

was displayed for 750 ms, followed by a delay of 500 ms, followed then by the target, which lasted until the participant's response. A 500 ms white noise mask was presented immediately after the target , followed by an inter-trial delay of 500 ms. Half of the participants indicated that the stimulus was concave by pressing the left button of the mouse with the right index finger and that the stimulus was convex with the right mouse button with the right middle finger. These assignments were reversed for the other half of participants. There were 30 trials per condition. Each stimulus was presented with each illumination direction and display mode seven or eight times, with the rule that the shapes would be presented at least seven times each, and that two of them were selected randomly to be presented an eighth time to equal 30 trials. This gave a total of 360 experimental trials divided in three blocks of 120 trials. In each block, there were 10 trials for each condition. The order of the blocks was random. The trials were presented in a random sequence within each block. Twenty practice trials were presented prior to the experimental trials. The dependent variable was the rate of "convex" responses.

Results

A three-way within-subject ANOVA including the factors of display mode (3D, reversed 3D, and 2D), illumination from above/below and illumination from left/right was carried out on the rates of "convex" judgements (Fig. 3.) Main effects of display mode (F (2, 26) = 16.81, $p < .001$, $\eta^2 = .56$) and illumination from above/below (F (1, 13) = 6.68, $p < .05$, $\eta^2 = .34$) were obtained. The display mode effect indicates that participants responded "convex" more often when the stimuli were shown with normal binocular disparity than with no disparity (F (1, 13) = 6.45, $p < .05$, $\eta^2 = .33$), and when the stimuli were presented with no disparity than with reversed disparity (F (1, 13) = 10.89, $p < .01$, $\eta^2 = .46$). The illumination above/below effect indicates that participants judged the stimuli as convex significantly more often when they were lighted from above than from below. We found no other significant effect (with all *F's* < 1). Indeed, the effect of illumination from left/right (*F* (1, 13) = .01, *p* = .95, η^2 = .001) was far from significance. Most importantly, the interaction of lighting from above/below x display mode (*F* (2, 26) = .23, *p* = .80, η^2 = .02.) was also far from significance thereby indicating the additivity of these factors. The effect of lighting from above/below was significant and of constant magnitude, regardless of whether the stimuli were presented with stereopsis (*F* (1, 13) = 6.22, *p* < .05, η^2 = .32, mean difference of .08 (SD = .12)), reversed stereopsis (*F* (1, 13) = 5.08, *p* < .05, η^2 = .28, mean difference of .08 (SD = .14)) or no disparity (*F* (1, 13) = 4.77, *p* < .05, η^2 = .27, mean difference of .10 (SD = .17).)

Insert Fig. 3 near here

Discussion

The results show that stereopsis and lighting from above/below both influence shape judgements. Thus, participants responded that the shapes were convex more often when the stimuli were presented with stereopsis than with no disparity, and when the stimuli were presented with no disparity than with reversed stereopsis. Also, participants responded "convex" more often when the lighting came from above than from below. However, whether the stimuli were lighted from the right or the left did not affect responses. The effects of display mode and lighting from above/below were precisely additive (*F* < 1).

The integration of depth cues

The absence of statistical interaction between the effects of display mode and lighting, and the fact that the effect sizes of the above/below lighting direction were almost exactly the same in the three display mode conditions (i.e. the difference across display modes is less than the standard deviation of the effect size for lighting direction) indicate that stereopsis and shading combine in an accumulative way. In other words they are integrated linearly. These findings are consistent with the MWF model (Landy et al., 1994) which postulates that depth cues are integrated linearly, as well as with the data of Dosher et al. (1986) which showed that stereo and proximity luminance covariance are integrated linearly.

However, our results may appear inconsistent with those of Bülthoff and Mallot (1988, 1990), who showed that edge-based stereo can override disparate shading and shape from shading in the evaluation of local surface depth. An attenuated or eliminated impact of above/below lighting direction on the perception of convexity/concavity could have been expected in the present results when stereoscopic information was available. We only found additivity of the effects, which may be due to a difference in the nature of the tasks required of the participants. Whereas we required a categorization of shapes as either convex or concave, the task used by Bülthoff and Mallot (1988, 1990) was one of local surface depth evaluation with objects that were either flat or that had variable degrees of convexity. We believe that the markedly different constraints on the task and the nature of the stimulus interpretation required is probably what led to the difference in how depth information was used. In fact, Bülthoff and Mallot (1988) admitted that edge-based stereo vetoing disparate shading and shape from shading could occur only locally (as in their task) and not in global perception. What has been found here is that shading and stereo are integrated linearly when it comes to perception of global shape relief.

Our results appear inconsistent with the studies showing that monocular cues may override stereopsis (Braunstein et al., 1986; Stevens & Brookes, 1988) since the direction of illumination information carried by shading did not alter the effect of display mode. It should be noted however, that the monocular depth cues studied in Braunstein et al. (1986) and in Stevens and Brookes (1988) are occlusion and velocity or texture gradients, in contrast to shading in the present study. This may thus suggest that different monocular depth cues differ in the degree to which they contribute to the interpretation of shapes in depth. A mitigating factor that may also be important in determining the contribution of different depth cues is the type of stimuli used. Thus, the stimuli used by Braunstein et al. and in Stevens and Brookes are line drawings, in contrast to the shaded objects used here.

It is important to note that in the normal stereo condition, when the shape was lighted from above, the convex response rate was of 91%, but this rate did not fall to 9% with the shape presented with reversed stereopsis and light from below. Indeed, in this condition, the convex response rate was of 32%. Thus, the results with reversed stereo are not the mirror image of those in the normal stereo condition. The convexity prior may be a factor in this asymmetry.

Perhaps more importantly, in the reversed stereo condition, the binocular disparities in the display were uncrossed whereas they were crossed in the normal stereo condition. Patterson et al. (1995) showed that perceived depth is more accurate and sensitive with crossed than with uncrossed disparities. However, their display durations were much shorter than ours (around 100 ms vs. 2732 ms). With their small stimulus duration it is impossible for participants to change their vergence angle while exploring the stimuli. In contrast, with the long stimulus exposures used here, vergence changes that alter the sign of binocular disparities may have occurred. Thus, it is not entirely clear whether the findings of Patterson et al. apply to account for the asymmetry observed here between normal and reversed stereo. Possibly more relevant, Tam and Stelmach (1998) found an asymmetry in stereoanomaly between crossed and uncrossed disparities. Indeed, they report that uncrossed disparity results in a greater number of participants failing to perceive stereoscopic depth than crossed disparity, and this with display durations as long as 1000 ms. In fact, at this stimulus duration, the difference between crossed and uncrossed disparities had long reached an asymptote, such that one should expect the same result with protracted stimulus exposures. This robust asymmetry between crossed and uncrossed disparities is congruent with the hypothesis that they are processed by different systems (Mustillo, 1985). Also supporting this view, Ishigushi and Wolfe (1993) demonstrated a difference in stereo capture between crossed and uncrossed disparities. Thus, while crossed disparity led to strong stereo capture, uncrossed disparity led to unstable representations. The authors accounted for this finding by suggesting that the two kinds of disparities play a different role in surface reconstruction and that they differ in their perceptual representation.

The asymmetry reported here between normal and reversed stereo thus agrees with previous relevant findings in the literature and with the accounts proposed by Mustillo (1985) and by Ishigushi and Wolfe (1993). Regardless of the specific reasons for this asymmetry, it remains that shading and stereo contribute independently to the perception of relief. In relation to this issue, it is interesting to note that since crossed and uncrossed disparities seem to be processed by different mechanisms (Mustillo, 1985) and that uncrossed disparity leads to unstable perception of depth (Ishigushi & Wolfe, 1993), one might have expected shading to have a greater impact with reversed than with normal stereo. This possibility was tested and rejected by the present study.

Lighting direction priors

Our results indicate that stereopsis does not affect the light-from-above prior. Thus, when stimuli were lighted from above, they were perceived as convex more often than when they were lighted from below, regardless of display mode. The effect of the above/below lighting direction observed here is consistent with the findings of Adams (2007), Connor (2001), Gerardin, de Montalembert and Mamassian (2007), and Kleffner and Ramachandran, (1992), who all showed that shape perception is constrained by the light-from above prior.

The left/right lighting direction had no effect on convexity judgements. This is in agreement with the results of Exp.1 by McManus et al. (2004), and with the notion that shape perception is not guided by a light-from-the-left prior. However, these results contradict those of Sun and Perona (1998) and of Mamassian and Goutcher (2001) as well as those of Exp. 2 by McManus et al., which suggested a lighting-from-the-left prior. It appears that exposure duration may be the factor responsible for these inconsistent results. In the present experiment, the stimuli remained visible until the participant’s response, as in Exp. 1 by McManus et al., which showed no leftward bias. The mean response time and thus, exposure duration, in the present experiment was of 2732 ms, and of 2068 ms in McManus et al. In contrast, exposure durations were shorter in the experiments that showed a leftward bias. Indeed, stimuli were presented for 120 ms in Mamassian and Goutcher, less than 500 ms in Sun and Perona and between 200 and 1000 ms in Exp. 2 by McManus et al. It is possible that longer exposure durations lead the participants to adopt a different response strategy since they have more time to analyse the available information. Sun and Perona found a strong correlation between handedness and the preferred lighting direction. This suggests that hemispheric laterality could be implicated in the leftward bias. Under this assumption, perhaps longer exposure durations leave more time for inter-hemispheric information transfer, which eliminates any potential lateralization effect that may otherwise occur with shorter durations. This could explain why short and long exposure durations are associated with different outcomes regarding the effect of left/right lighting direction.

Another feature of the present results is the relative magnitude of the effects of display mode and of shading on the frequency of “convex” responses. Overall, this frequency is increased by 20% with normal stereo relative to 2D presentations and decreased by 30% with reversed stereo in comparison to 2D displays. These effects are of a much greater magnitude than the overall impact of above/below lighting direction, which differed by 9% in the rate of “convex” responses. This suggests that in the present experimental context, stereoscopic information carried a greater weight for the interpretation of shapes in depth than shading. This finding is congruent with the notion of the relative dominance of stereo over shading for the perception of 3D shapes that we may retain from the studies of Bülthoff and Mallot (1988, 1990) discussed above.

It is possible however, that the relative importance of different depth cues varies according to the strength of the signal they offer. For instance, had we used the same object as in the present study but with a greater simulated distance, binocular disparities would have been smaller and this could have reduced or even eliminated the dominance of stereo over above/below lighting. Alternatively, maybe the same effect could be obtained with a manipulation of above/below lighting that involves greater position disparities between the two light sources. These issues will need further studies to be elucidated.

Convexity prior

We may note also that with 2D displays, participants responded "convex" on more than 50% of trials even with lighting from below. We interpret this finding as further support for the observations of Langer and Bülthoff (2001) demonstrating a bias towards an interpretation of shapes as convex. Furthermore, these findings suggest, in support of previous results by Liu and Todd (2004), that the bias for convexity is stronger than the bias for the above/below direction of illumination.

It is also important to note that in the normal stereo condition, when the stimuli were lighted from above, the rate of convex responses was of 90% instead of 100%, as we might have expected. We note however, that the depth information carried by the stimuli was rather impoverished. Indeed, the only available depth cues were binocular disparity, whose informativeness was probably advantaged by the flat extremity on the shapes, and shading. It is possible that with so little information, some ambiguity remained about the convexity of the stimuli, even with normal stereo and lighting from above.

Summary

The goal of the present research was to determine if shading and stereopsis have independent or interactive contributions to shape perception. An experiment using a shape judgement task (i.e. determine if the shape is convex or concave) was designed. The results show that the effects of display mode and lighting direction are additive; i.e. they do not interact. This indicates that stereopsis and shading have their own independent contributions to shape perception. In other words, each depth cue appears to be processed independently and to affect convexity/concavity judgements to a degree that is independent of the direction suggested by the other cue.

Acknowledgements

The research reported here was supported by grant no. 203613 from the Natural Sciences and Engineering Research Council of Canada to Martin Arguin.

References

Adams, W. J. (2007) A common light-prior for visual search, shape, and reflectance judgements, *Journal of Vision*, *7*(11), 1-7.

Braunstein, M. L., Andersen, G. J., Rouse, M. W., & Tittle, J. S. (1986) Recovering viewer-centered depth from disparity, occlusion and velocity gradients, *Perception & Psychophysics*, *40*(4), 216–224.

Brewster, D. (1826) On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of another analogous phenomena, *Edinburgh Journal of Science*, *4*(7), 99-108.

Bülthoff, H. H., & Mallot, H. A. (1988) Integration of depth modules: stereo and shading, *Journal of Optical Society of America*, *5*(10), 1749-1758.

Bülthoff, H. H., & Mallot, H. A. (1990) Integration of stereo, shading and texture. In A. Blake, & T. Troscianko (eds.), *AI and the Eye* (pp.119-146). Toronto: Wiley

Champion, R. A., & Adams, W. J. (2007) Modification of the convexity prior but not the light-from-above prior in visual search with shaded objects, *Journal of Vision*, *7*(13), 1-10.

Connor, C.E. (2001) Visual Perception: Sunny side up, Current Biology, *11(19)*, R776- R778.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986) Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure, *Vision research*, *26*(6), 973-990.

Gerardin, P., de Montalembert, M., & Mamassian, P. (2007) Shape from shading: New perspectives from the Polo Mint stimulus, *Journal of Vision*, *7*(11), 1-11.

Hill, H., & Bruce, V. (1994) A comparison between the hollow-face and 'hollow-potato' illusions, *Perception*, *23*(11), 1335-1337.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004) Slant from texture and disparity cues: optimal cue combination, *Journal of Vision*, *4*(12), 967-992.

Ichikawa, M., & Egusa, H. (1993) How is depth perception affected by long-term wearing of left-right reversing spactacles?, *Perception*, *22*(8), 971-984.

Ichikawa, M., Egusa, H., Nakatsuka, M., Amano, J., Ueda, T., & Tashiro, T. (2003) Modification of depth and distance perception caused by long-term wearing of left-right reversing spectacles, *Perception*, 32(2), 131-153.

Ishigushi, A., & Wolfe, J. M. (1993) Asymetrical effect of crossed and uncrossed disparity on stereoscopic capture, *Perception*, *22*(12), 1403-1413.

Kleffner, D. A., & Ramachandran, V. S. (1992) On the perception of shape from shading, *Perception and Psychophysics*, *52*(1), 18-36.

Knill, D. C., & Saunders, J. A. (2003) Do human optimally integrate stereo and texture information for judgement of surface slants?, *Vision Research*, *43*(24), 2536-2558.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1994) Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research*, *35*(3), 389-412.

Langer, M. S., & Bülthoff, H. H. (2000) Depth discrimination from shading under diffuse lighting, *Perception, 29*(6), 649-660.

Langer, M. S., & Bülthoff, H. H. (2001) A prior for global convexity in local shape from shading, *Perception*, *30*(4), 403-410.

Liu, B., & Todd, J. T. (2004) Perceptual biases in the interpretation of 3D shape from shading, *Vision Research*, *44*(18), 2135-2145.

Liu, Y., Vogels, R., & Orban, G.A. (2004) Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex, *The Journal of Neuroscience, 24(15),* 3795-3800.

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

Mamassian, P., & Landy, M. S. (1998) Observer biases in the 3D interpretation of line drawings, *Vision Research*, *38*(18), 2817-2832.

Mamassian, P., & Landy, M. S. (2001) Interaction of visual prior constraints, *Vision Research*, *41*(20), 2653-2668.

Matthews, H., Hill, H., & Palmisano, S. (2011) Binocular disparity magnitude affects perceived depth magnitude despite inversion of depth order, *Perception*, *40*(8), 975-988.

McManus, I. C., Buckman, J., & Woolley, E. (2004) Is light in pictures presumed to come from the left side?, *Perception*, *33*(12), 1421-1436.

Mustillo, P. (1985) Binocular mechanisms mediating crossed and uncrossed stereopsis, *Pyschological Bulletin*, *97*(2), 187-201.

Norman, J. F., & Todd, J. T. (1995) The perception of 3-D structure from contradictory optical patterns, *Perception & Psychophysics*, *57*(6), 826-834.

Patterson, R., Cayko, R., Short, L., Flanagan, R., Moe, L., Taylor, E., & Day, P. (1995) Temporal differences between crossed and uncrossed stereoscopic mechanisms, *Perception and Psychophysics*, *57*(6), 891-897.

Ramachandran, V. S. (1988) Perception of shape from shading, *Nature*, *331*(6152), 163-166.

Reichel, F. D., & Todd, J. T. (1990) Perceived depth inversion of smoothly curved surface due to image orientation, *Journal of Experimental Psychology: human perception and performance*, *16*(3), 653-664.

Saunders, J. A., & Backus, B. T. (2006) Perception of surface slant from oriented textures, *Journal of Vision, 6*(9), 882-897.

Shimojo, S., & Nakajama, Y. (1981) Adaptation to the reversal of binocular depth cues: Effects of wearing left-right reversing spectacles on stereoscopic depth perception, *Perception, 10*(4), 391-402.

Stevens, K. A., & Brookes, A. (1988) Integrating stereopsis with monocular interpretations of planar surfaces, *Vision Research, 28*(3), 371-386.

Sun, J., & Perona, P. (1998) Where is the sun?, *Nature Neuroscience, 1*(3), 183-184.

Tam, W. J., & Stelmach, L. B. (1998) Display duration and stereoscopic depth discrimination, *Canadian Journal of Experimental Psychology, 52*(1), 56 -61.

Thomas, R., Nardini, M., & Mareschal, D. (2010) Interaction between “light-from-above” and convexity priors in visual development, *Journal of Vision, 10*(8), 1-7.

Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002) Neural correlates for perception of 3D surface orientation from texture gradient, *Science, 298(5592)*, 409-412.

Yellott, J. I., & Kaiwi, J. L. (1979) Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective, *Perception, 8*(2), 135-142.

Figure captions

FIG. 1: The four views under which the shape used in the present experiment could be displayed. The different viewpoints were created by rotating the object by 90 deg around the z-axis relative to one another.

FIG. 2: The four lighting directions used (i.e. above-left, above-right, below-left, below-right) illustrated with the top-left object of Fig. 1.

FIG. 3: The rates of “convex” responses as a function of the above/below and left/right directions of illumination and of display mode.

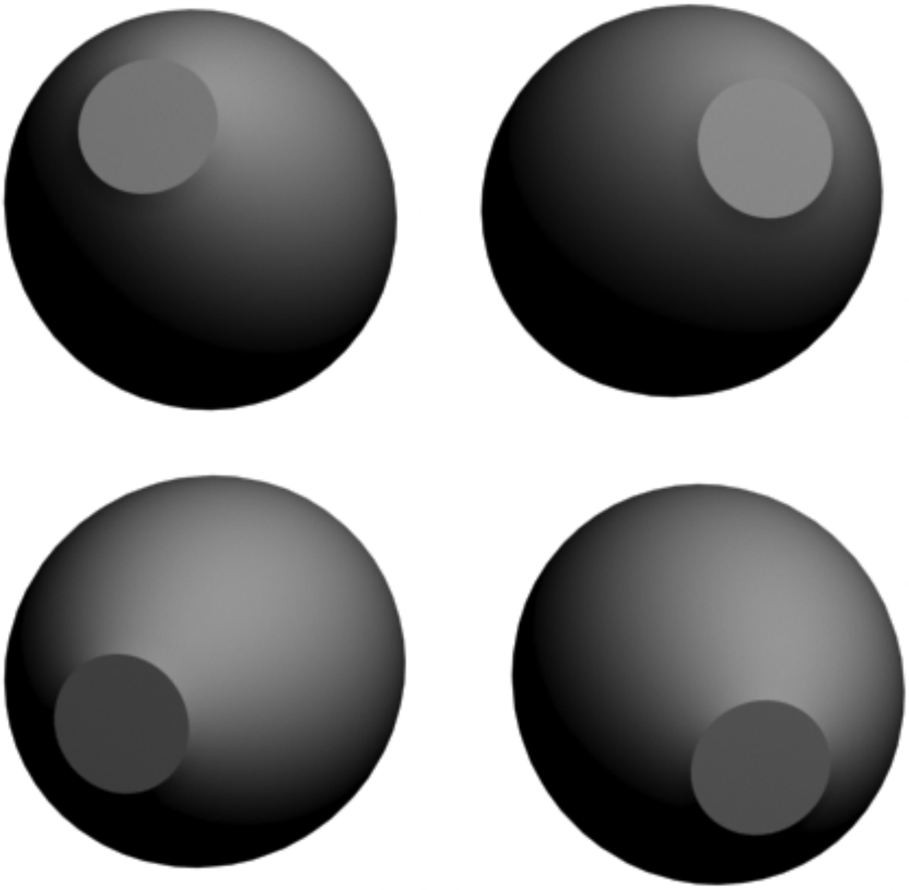


Fig. 1

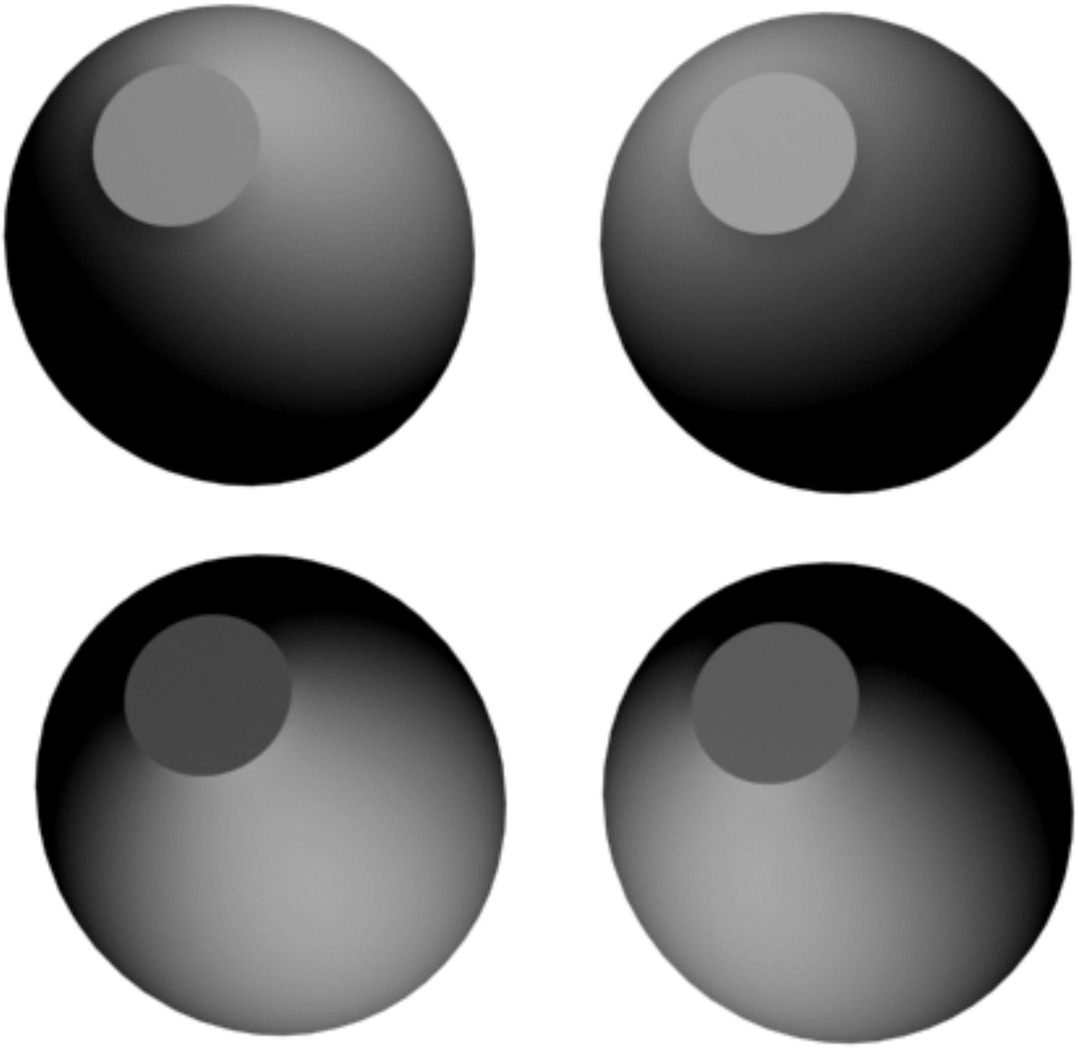


Fig. 2

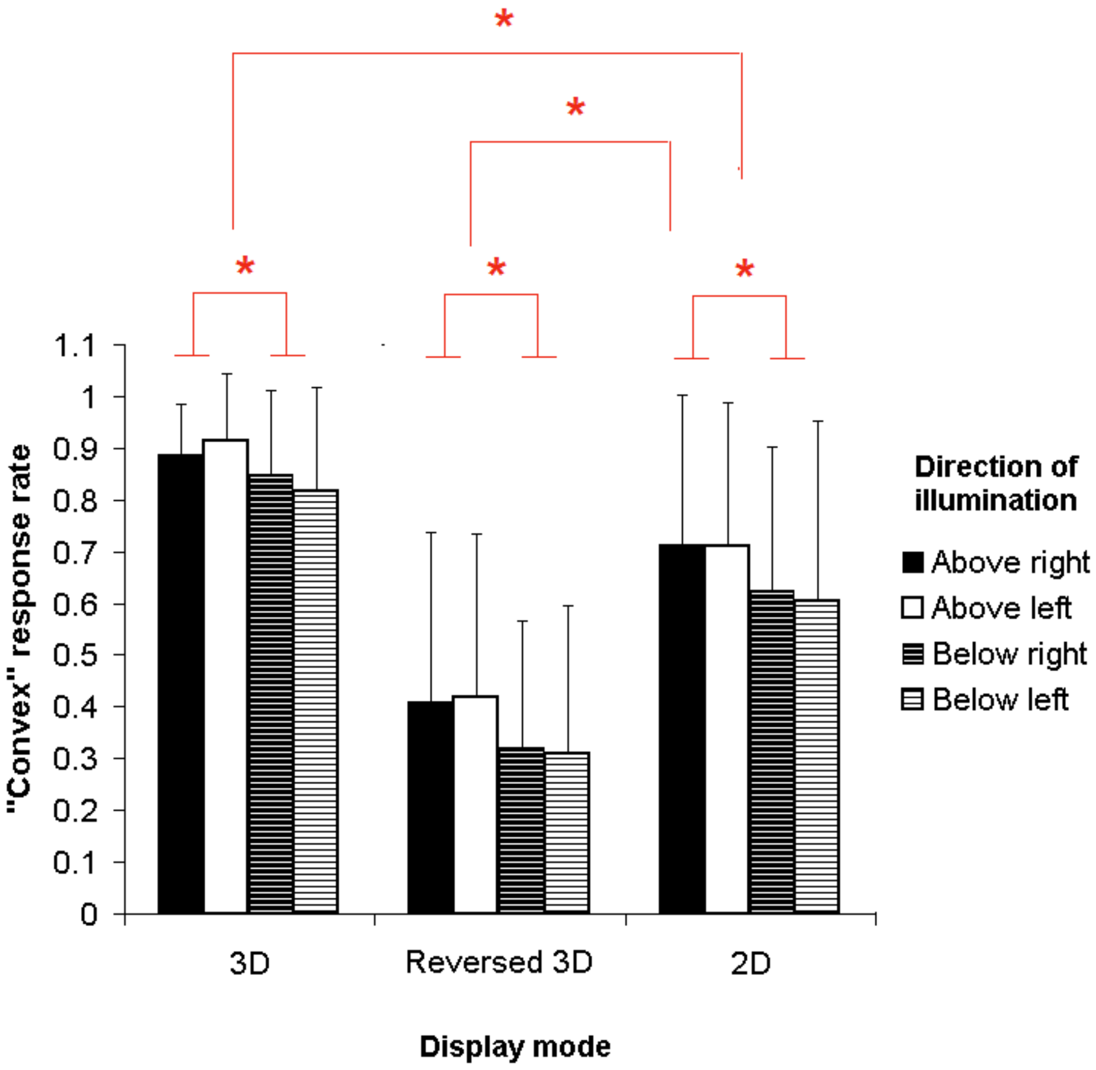

*
*
*
*
*
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
"Convex" response rate
3D
Reversed 3D
2D
Display mode
Direction of illumination
■ Above right
□ Above left
▤ Below right
▤ Below left

Fig. 3

Article 3

THE CONDITIONAL CONTRIBUTION OF STEREOPSIS TO SHAPE CONSTANCY

Running head: Stereopsis and shape constancy

Mercédès Aubin, Sacha Zahabi & Martin Arguin

Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de
psychologie, Université de Montréal

Contribution des coauteurs :

M. Aubin., & M. Arguin. : Conception du projet

M. Aubin. & S. Zahabi : Recrutement des participants

M. Aubin. & S. Zahabi : Collecte des données

M. Aubin. : Analyse des résultats

M. Aubin. : Rédaction de l'article

M. Aubin. : Interprétation des résultats

M. Aubin, & M. Arguin : Révision du manuscrit

M. Arguin : Financement

Abstract

We assessed the contribution of stereopsis to shape constancy and how it interacts with perceptual expertise using sequential matching tasks with bent paperclips rotated in depth. Stimuli were presented without stereopsis (same view presented at each eye), or with normal or reversed stereopsis. In the first half of Exp. 1, where display mode was within-subject, the rotation effect was weaker with stereoscopic displays than with 2D or reversed stereoscopic presentations. These effects were reversed in the second half of the experiment. In Exps. 2 (display mode within-subject, with stimuli switched halfway into the experiment) and 3, (display mode between-subjects) the rotation effect was consistently weaker with normal stereo than with either 2D or reversed stereoscopic displays. These experiments also demonstrate an advantage of reversed stereo over 2D presentations. We conclude that stereo may contribute to shape constancy. This, however, requires stereoscopic information validity to vary minimally across the trials. Otherwise, stereo may become strategically ignored.

Keywords: shape constancy, stereopsis, shape recognition, shape perception, perceptual learning

Understanding the representation of three-dimensional (3-D) objects is fundamental for theories of vision (e.g., Arguin & Saumier, 2004; Biederman, 1987; Edelman, 1999; Hummel, 2001; Hummel & Biederman, 1992; Leek, Reppa, & Arguin, 2005; Marr & Nishihara, 1978). One unresolved issue concerns the contribution of the stereoscopic depth cue for recognition and shape constancy, and its relation with perceptual expertise.

According to current structural theories (Biederman, 1987; Biederman & Bar, 1999; Biederman & Gerhardstein, 1993; Hummel & Biederman, 1992; Marr, 1982; Marr & Nishihara, 1978; Peissig, Wasserman, Young, & Biederman, 2002; Pentland, 1985), internal shape representations encode the 3D structure of objects in terms of a collection of volumetric primitives and their connectedness. Biederman proposed that the non-accidental properties, which are 3D shape properties that are shared across the viewpoints of an object, would be the basis for its recognition. Thus, according to this view, as long as the non-accidental properties remain visible, shape representations should be orientation invariant. As reported by Biederman, the implicit 3D cues provided by the processing of edges are sufficient to create a 3D shape representation allowing shape constancy. Other kinds of depth cues like texture, shading or stereopsis are assumed not to contribute. Congruently, Biederman & Ju (1988) showed that object naming was just as fast for line drawings as for photographs.

View-based theories, which constitute another major class of theories, suggest that shape representations exclusively code information about the 2D views. This necessarily implies that the recognition of an object in a novel orientation requires a special process of alignment, normalisation and/or interpolation to match it to its stored 2D representation (Hayward & Tarr, 1997; Tarr & Bülthoff, 1995; Tarr, Bülthoff, Zabinski, & Blanz, 1997; Tarr & Pinker, 1989, 1990, 1991; Tarr, Williams, Hayward, & Gauthier, 1998). Proponents of view-based theories have proposed that depth cues, including stereopsis, may contribute to shape perception, but only in a view specific manner – i.e. with no contribution to shape constancy (Edelman and Bulthöff, 1992).

Proposing a different theory, Pizlo and colleagues (Chan, Stevenson, Li, & Pizlo, 2006; Pizlo, 2008; Pizlo, Li & Francis, 2005; Pizlo, Li, Steinman, 2008; Pizlo & Stevenson, 1999) argue that shape representation and shape constancy are based on a priori monocular simplicity constraints. Consequently, binocular depth cues would be neither necessary nor sufficient for shape constancy. In support of the latter, Pizlo et al. (2005) have reported that when monocular simplicity constraints and stereoscopic depth cues are inconsistent, binocular cues are ignored. Furthermore, the performance of models of monocular shape representation in a sequential discrimination task correlate with human monocular and binocular performances while the performances of binocular models do not (Chan & al, 2006).

Despite the results presented above, other evidence indicates that stereopsis reduces rotation costs in different matching and recognition tasks with stimulus classes for which humans fail to show orientation invariance, i.e. faces, bent paper clips, and other kinds of tubular shapes (Bennett & Vuong, 2006; Burke, 2005; Burke, Taubert, & Higman, 2007). Conversely, Pasqualotto and Hayward (2009) showed that under particular task conditions, stereopsis may alter recognition performance with rotated familiar objects. The authors have suggested that stereoscopic cues may reduce the weight of general similarities between the shapes in the recognition process, thereby increasing the weight of fine and precise information. According to the authors, it is this extra-specificity in processing that led to reduced performance in the stereoscopic condition. Whether this interpretation is correct or not, the results of Pasqualotto and Hayward support the notion that stereo is involved in shape perception.

Palmeri, Wong and Gauthier (2004) discuss changes in the processing of visual objects that accompany the development of perceptual expertise. They underline that a shift from relying on explicit rules to relying on encoded representations, an increase in memory sensitivity, an increase in perceptual sensitivity, changes in selective attention, and changes in perceptual representation may accompany the development of perceptual expertise. The changes in selective attention may lead the observer to attend selectively to the object dimensions that are the most informative for making the discriminations required by the task. It is interesting to examine the results of Pasqualotto and Hayward (2009) in this light. Indeed, these authors used familiar objects, for which the observers had necessarily developed some sort of perceptual expertise. From their findings, it seems that stereo might reduce the benefit of expertise by forcing finer discriminations instead of attending maximally to the most informative elements of the stimuli. Given that Pasqualotto and Hayward used objects that were already familiar, it is impossible to tell if the deleterious effect of stereo actually grew out of the development of expertise with the objects used. Only by using objects that are initially unfamiliar would it be possible to determine the change in the impact of stereoscopic information in shape discrimination which may occur with the development of expertise.

In another vein, to explore the contribution of stereo to shape perception, reversed stereo appears as a relevant exposure condition that has been poorly explored so far. Thus, the studies that have examined the effect of reversed stereopsis on visual perception (Ichikawa & Egusa, 1993; Ichikawa, Egusa, Nakatsuka, Amano, Ueda, & Tashiro, 2003; Shimojo & Nakajama, 1981; Wheatsone, 1852; Yellot & Kaiwi, 1979) have focused on the adaptation effect resulting from the long term wearing of right-left reversing spectacles which reversed binocular disparities. This adaptation led to depth inversion once the spectacles were removed, and it also altered the weight of the different depth cues (Ichikawa & Egusa, 1993). Indeed, participants ended up ignoring binocular disparity altogether and using occlusion and linear perspective to a greater degree to make depth judgements than before the adaptation.

No study has yet been conducted to determine the impact of reversed stereopsis on shape constancy. This is relevant, because if the visual system uses stereopsis to achieve shape constancy, we would expect reversed disparity to alter performances compared to normal disparity or the absence of disparity since stereo would be inconsistent with others depth cues like shading, edges, occlusion, etc. However, the information brought by reversed disparity could also be useful to achieve rotation invariance. Indeed, even though its sign is inconsistent with the other depth cues, its magnitude is not. If useful because of the valid depth magnitude information it provides, reversed 3D should benefit performance and reduce rotation costs (i.e. increase of the response times and error rates as a function of the amount of rotation between two stimuli) compared to 2D. Thus, this condition could inform us further about the role of stereopsis in shape constancy and about how it is used.

The goals of the present study are to determine the contribution of stereopsis to shape constancy in the visual perception of non-familiar shapes which do not respond to the conditions for orientation invariance proposed by Biederman & Gerhardstein (1993) and to evaluate the impact of the development of expertise on the contribution of stereopsis. The experiments had a methodology quite similar to that of Burke (2005), who used a sequential matching task with bent paper-clips as stimuli. A display mode of reversed disparity was added to the normal binocular disparity and no disparity between the views presented at each eye conditions of Burke (2005), and we had more trials per participant in order to compare the performance in the first and second halves of the experiments. The first experiment closely replicated that of Burke, with the addition of a condition of reversed binocular disparity to those of normal binocular disparity and no disparity between the views presented at each eye. As in Burke (2005), these display conditions were within-subject. Exp. 2 replicated Exp. 1 but with a switch of stimuli in the course of the experiment to examine the evolution of processing when the development of object specific perceptual expertise is interrupted. Finally, Exp. 3 replicated Exp. 1 except that the different display modes were between-subjects, which allowed us to evaluate the development of expertise with only one display mode

Experiment 1

Exp. 1 used a sequential matching task with rotated bent paper-clips that were displayed in one of three display modes (stereo, reversed stereo, or no stereo). Display mode varied in a within subject manner. It was expected that rotation costs would be lower with stereoscopic displays than with either reversed stereo or null disparity.

Methods

Participants. Twelve (1 males and 11 females, 11 right-handers and 1 left-hander) participants aged between 18 and 30 years old took part in the experiment. All were naive as to the purpose of the experiment, were neurologically intact, and had normal or corrected visual acuity. No particular selection was applied with respect to gender, manual dominance, or level of education.

Material and Stimuli. The stimuli were presented on a 16-inch Compaq monitor of 1024 x 768 pixels resolution. The experiment was run on a Pentium 4 computer, and its progress and the registration of the observer's responses were controlled by the E-Prime software. Participants responded by pressing the buttons of a Serial Response Box by Psychology Software Tools.

The stimuli are shown in Fig. 1. They were synthetic renditions of Burke's (2005) stimuli (which were photographs of actual bent paperclips) created using the 3D Studio MAX® program from Autodesk Media & Entertainment (CA, USA). These were four bent paperclip-like shapes with a spatial extent of 28 x 33 degrees of visual angle presented at a viewing distance of 35 cm. Each paperclip was presented at five possible viewpoints (0, 20, 40, 60 or 80 degrees of rotation around the vertical axis.) Rotations of 10.6 degrees around the vertical axis were applied to the stimuli to create distinct views for the left and right eyes for the stereo and reversed stereo conditions. The images were then fused to produce red/cyan anaglyphs. For the 2D display condition, stimuli with no disparity were fused to make their preparation and exposure identical across conditions (all stimuli were viewed with red/cyan glasses, which participants wore throughout the experiment).

Insert Fig. 1 near here

Procedure. Participants performed a sequential matching task with a 3 (display mode: 3D, reversed 3D or 2D) x 2 (same or different) x 5 (orientation: 0^o, 20^o, 40^o, 60^o, 80^o) repeated measures experimental design. The viewing distance of 35 cm was maintained constant using a chin rest. The background of the screen was neutral grey. At the beginning of each trial, a fixation cross was displayed for 500 ms, followed by the first stimulus for 2500 ms, then by a white noise mask for 500 ms, and finally by the second stimulus, which lasted until the participant's response or up to a 5000 ms delay. Visual feedback ("incorrect" written in the middle of the screen) of 1000 ms duration was presented when the participant made an error. There was a 500 ms delay between each trial. Participants were instructed to respond as fast and accurately as possible. Half of the participants indicated that the stimuli were the same by pressing the right button of the response box with the right index finger and that the stimuli were different with the left button with the left index finger. These assignments were reversed for the other half of participants.

For each object, we can construct 25 "same" pairs by varying the orientation of the stimuli and their order (e.g. Stimulus 1 at 0^o and Stimulus 1 at 0^o, Stimulus 1 at 0^o and Stimulus 1 at 20^o, and so on). This gives a total of 100 "same" combinations across the four objects. Each of these combinations was presented twice for each display mode. This gives a total of 200 "same" trials for each display mode. There were also 200 "different" trials for each display mode. They had the same structure as "same" trials, but the second stimulus in the sequence was replaced by a randomly selected different object which was shown from the same orientation as the one it replaced. The complete experiment thus comprised a grand total of 1200 trials, which were divided in two sessions of three blocks of 200 trials each. All the 100 possible combinations of "same" stimuli for each display mode were presented within each session and distributed randomly across blocks. The "different" trials were distributed randomly with the constraint that each object, orientation and display mode occurred in equal numbers in each session. Within each block, there were 100 "same" trials and 100 "different" trials in a random sequence.

Results

Response times (RTs) that were more than 2.5 standard deviations away from a participant's average for a given condition were eliminated, resulting in the exclusion of a total of 315 data points for the complete experiment (2.2% of trials). The error trials were also excluded from the RTs analyses (11.0% of trials). The data showed no speed-accuracy trade-off since the correlation between correct RTs and error rates was positive, $r = .77$, $p < .05$.

RTs analysis. Data analyses were conducted for "same" trials only since the notion of orientation difference is irrelevant for different objects. Fig. 2 displays the average correct RTs as a function of the degree of rotation for the two sessions. Table 1 presents the results of the linear regression analyses of RTs as a function of rotation. The regression slopes help to illustrate the amplitude of the rotation costs. Indeed, the higher are the slopes and the higher are the rotation costs.

Insert Fig. 2 and Table 1 near here

A three-way within-subject ANOVA including the factors of training (session 1 and 2), rotation (0, 20, 40, 60 and 80 degrees of rotation) and display mode (2D, stereo, reversed stereo) was carried out on correct RTs. Main effects of training, $F(1, 11) = 43.75$, $p < .001$, rotation, $F(4, 44) = 27.98$, $p < .001$, and display mode, $F(2, 22) = 8.19$, $p < .01$, were obtained. The effects of training and rotation indicate that RTs diminished from session 1 to session 2, and that they increased with rotation. The main effect of display mode indicates that RTs were shorter for the reversed 3D, $F(1, 11) = 14.44$, $p < .01$, and the 2D, $F(1, 11) = 8.98$, $p <.05$, than for the 3D conditions. However, there was no significant RT difference between the 2D and the reversed 3D conditions, F (1, 11) < 1. These main effects were qualified by a significant three-way interaction of training x rotation x display, F (8, 88) = 3.04, $p < .01$. This result indicates that the impact of display mode on the rotation effect was modulated by training.

In the first session, there were significant effects of rotation, F (4, 44) = 23.31, $p < .001$, and of display mode, F (2, 22) = 33.55, $p < .001$, but no interaction between these factors, F (8, 88) = 1.05, *ns*. RTs were significantly shorter in the 3D display mode than in the 2D, F (1, 11) = 50.34, $p < .001$, or the reversed 3D conditions, F (1, 11) = 30.99, $p < .001$. The RTs in the latter conditions did not differ, F (1, 11) = 2.37, *ns*.

In the second session, there were significant effects of rotation, F (4, 44) = 21.37, $p < .001$, and of display mode, F (2, 22) = 54.64, $p < .001$, and the interaction between these factors was also significant, F (8, 88) = 3.73, $p < .001$. RTs were significantly longer in the 3D display mode than in the 2D, F (1, 11) = 57.97, $p < .001$, or the reversed 3D conditions, F (1, 11) = 56.88, $p < .001$. There was no significant difference between the 2D and the reversed 3D conditions, F (1, 11) = 1.57, *ns*. Pairwise contrasts between display mode conditions of the second session indicated that the rotation cost was greater in the 3D condition than with either the 2D, F (4, 44) = 7.56, $p < .01$, or reversed 3D displays, F (4, 44) = 3.24, $p < .05$ (see also the slopes in Table 1). The contrast between the 2D and reversed 3D conditions showed no significant difference in rotation costs, F (4, 44) = 1.02, *ns*.

Error rates analysis. Fig. 3 displays the average error rates (ERs) as a function of the degree of rotation for the two sessions. Table 1 presents the results of the linear regression analyses of ERs as a function of rotation for the two sessions.

Insert Fig. 3 near here

A three-way within-subject ANOVA including the factors of training (session 1 and session 2), rotation (0, 20, 40, 60 and 80 degrees of rotation) and display mode (2D, stereo, reversed stereo) was carried out on the ERs. A main effect of training was obtained, *F* (1, 11) = 6.98, *p* < .05, which indicated lower ERs in session 2 than in session 1. The main effect of display mode was also significant, *F* (2, 22) = 5.11, *p* < .05. ERs did not differ between the 2D and 3D display modes, *F* (1, 11) < 1, but they were significantly lower in the reversed 3D than in either the 3D, *F* (1, 11) = 15.48, *p* < .01, or the 2D conditions, *F* (1, 11) = 5.23, *p* < .05. The main effect of rotation was also significant, *F* (4, 44) = 26.36, *p* < .001, which indicated a significant increase of ERs as a function of rotation. The two-way interaction of training x rotation was significant, *F* (4, 44) = 2.70, *p* < .05. This result indicates that the rotation effect was reduced by training, as illustrated in Fig. 3 and by the slopes in Table 1. The three-way interaction of training x rotation x display mode was marginally significant, *F* (2, 22) = 2.81, *p* = .08. As can be seen in Fig. 3, this trend for an interaction consists in a reversal of the ERs pattern as a function of display mode across the two sessions. Even though none of the following contrasts are significant (all *p's* > .14) rotation costs tended to be weaker in the 3D and reversed 3D conditions than in the 2D condition in the first session, and weaker in the reversed 3D and 2D conditions than in the 3D condition in the second session.

Discussion

Globally, we found that performances improved from session 1 to session 2 and that they degraded with increasing rotation across the stimuli to be matched. Display conditions also had an important impact on performances and interacted with training and rotation. In particular, in the first session, RTs were shorter in the 3D than in the 2D or reversed 3D conditions. In contrast, in the second session, RTs as well as rotation costs were greater in the 3D than in the 2D or reversed 3D conditions. No significant difference was observed between the 2D and the reversed 3D display conditions.

The rotation costs observed in Exp. 1 were expected. Indeed, any of the theories described in the introduction predicts these effects since bent paperclips do not meet the rotation invariance conditions of Biederman and Gerharstein (1993) or the simplicity constraints of Pizlo and Stevenson (1999). Obviously, view-based theories also predicted the rotation costs of Exp. 1.

However, the display mode effect and its interactions with rotation and training could not be expected according to the results of Burke (2005). As Burke had found, we expected that rotation costs would be reduced by normal stereopsis (i.e. the 3D condition). In the first session, we did indeed find a performance advantage in the 3D condition, with reduced RTs compared to the 2D and reversed 3D conditions. However, in session 1, in contrast to Burke (2005), display mode failed to interact with rotation. Most unexpectedly, in the second part of the experiment, we found that rotation costs were *greater* in the 3D condition than with 2D or reversed 3D displays.

The hypothesis we propose to account for this finding is that it may result from the trial-to-trial inconsistency in the binocular disparity information to which participants were exposed. Specifically, we suggest that participants started the experiment with the strategy of attempting to construct a global 3D representation of the shapes using all the information available. This led to an RT advantage for the 3D condition in the first session since this condition offered more valid information than the others on the structure of the objects. However, in the 2D condition, no stereoscopic information was available and in the reversed 3D condition, this information was opposite to that offered by monocular depth cues. This inconsistency in the information offered by binocular disparity may have led participants to eventually focus on rotation-resistant 2D cues that they may have noted through exposure and which were available in all display conditions. This would be consistent with the idea that with the development of perceptual expertise, there can be a change in the focus of selective attention towards the most informative dimensions within the stimulus for the task that has to be performed (Palmeri et al., 2004). In the present case, with training, participants may have started to attend mostly to 2D rotation-resistant cues, which may have been quite sufficient to perform the task and which may be more quickly or easily available than stereo to signal the 3D structure of the objects.

This hypothesis however, does not fully explain why performance actually became worse in the 3D condition than in the other conditions in the second half of the experiment. In this regard, it is worth noting that Pasqualotto and Hayward (2009) have shown that particular test conditions may lead to a stereo disadvantage in the recognition of familiar objects. For the particular context of their experiment, they proposed that participants may have relied on the 2D outline shape to perform the recognition task and that “the extra specificity provided by stereoscopic disparity seems to have obscured more general similarities between the pictorial elements of the images, particularly the outline shapes” (Pasqualotto & Hayward, 2009; p.

837). It is possible here that stereo diverted attention away from the more effective 2D rotation-resistant cues, which would explain the longer RTs and greater ERs in the 3D condition in the second session. Thus, the development of perceptual expertise may have transformed stereo from a beneficial source of information early in the experiment, into an impediment later on. In contrast, reversed stereo may have been easier to reject, thereby leading to less interference, given that the depth information it offers is incongruent with the remainder of the information available in the stimulus.

In his study, Burke (2005) reported no inversion of the effect of display type. However, Burke's experiment comprised a total of 400 trials which were conducted over a single session, instead of the 1,200 trials administered over two sessions used here, which may have led our participants to develop a greater degree of expertise than those of Burke. Perhaps as importantly, Burke did not use reversed 3D displays, so participants were never exposed to stimuli in which the depth information provided by stereo was opposite that available from monocular depth cues. This may be another factor that is responsible for the performance cost associated with the stereo condition in the second half of our experiment. The other studies that have examined the role of stereopsis in shape constancy (Bennett & Vuong, 2006; Burke, Taubert, & Higman, 2007) also did not report an inversion of the stereo effect, but none of them comprised a reversed 3D condition either.

Exps. 2 and 3 were designed to assess the account proposed above to explain why normal stereo interfered with performance in the second half of Exp. 1 after having been beneficial in the first half. Exp. 2 is designed to assess whether object-specific expertise is involved in the reversal of the effect of normal stereoscopic exposure across the first and second halves of Exp. 1. Thus, Exp. 2 replicates Exp. 1 exactly, except for the fact that a new set of stimuli is introduced in the second half of the experiment rather than keeping the same set throughout the experiment. If, as argued above, the participants of Exp. 1 learned 2D cues which allowed them to perform the task and thus reject the information provided by normal stereo in the second half of the experiment, switching stimulus sets between sessions 1 and 2 in Exp. 2 will prevent such expertise to come into play and performance benefits from the 3D display mode should be maintained even in session 2. Exp. 3 will determine the impact of exposing participants to inconsistent depth information by randomly mixing trials using stimuli with normal, reversed, or no disparity between the views presented at each eye. Thus, instead of having display mode as a within-subject factor as in Exp. 1, Exp. 3 will use it as a between-subjects factor. By receiving consistently valid depth information, it is expected that participants in the normal 3D condition should show a performance advantage over participants receiving 2D or reversed 3D displays even in the second test session.

Experiment 2

The aim of Exp. 2 was to characterise the changes produced by perceptual expertise we found in Exp. 1. We have proposed that, through exposure, participants in Exp. 1 learned 2D cues that were relatively resistant to depth rotation. However, if the stimuli used in session 1 were to be replaced by new ones in session 2, the 2D cues learned in session 1 would become useless. Exp. 2 will use this manipulation to verify that perceptual expertise was indeed a factor in the evolution of the effect of normal stereo in the course of Exp. 1. Specifically, Exp. 2 used the same methods as Exp. 1 except for a change of stimuli between sessions 1 and 2.

Methods

Participants. Twelve participants (5 males and 7 females, 10 right-handers and 2 left-handers) between 18 and 39 years old took part in the experiment. All were naive as to the purpose of the experiment, were neurologically intact, and had normal or corrected visual acuity. No particular selection was applied with respect to gender, manual dominance, or level of education.

Stimuli. The stimuli used in Exp. 1 served again in Exp.2, along with a new set to allow a stimulus switch midway into the experiment. The new stimuli (see figure 4) were matched to the original set in terms of their size in the three dimensions. They have the same screen size as the original ones, and the anaglyphs were created in the same way.

Insert Fig. 4 near here

Procedure. The method is the same as for Exp. 1, except that there was a change of stimuli between sessions 1 and 2. Half of the participants started the experiment with the original set of stimuli used in Exp. 1 and the other half started it with the new set developed for the purpose of Exp. 2.

Results

RTs that were more than 2.5 standard deviations away from a participant's average for a given condition were eliminated, resulting in the exclusion of a total of 338 data points for the complete experiment (2.4% of trials). The error trials were also excluded from the RTs analyses (17.0% of trials). The data showed no speed-accuracy trade-off since the correlation between correct RTs and ERs was positive and significant, $r = .57$, $p < .05$.

RTs analysis. Figure 5 displays the average correct RTs as a function of the degree of rotation of the stimuli for the two sessions. Table 2 presents the results of the linear regression analyses of RTs as a function of rotation for the two sessions. A three-way within-subject ANOVA including the factors of training (sessions 1 and 2), rotation (0, 20, 40, 60, and 80 degrees of rotation) and display mode (2D, 3D, and reversed 3D) was carried out on correct RTs. The analysis showed only a significant main effect of rotation, $F(4, 44) = 43.80$, $p < .001$, (display mode: $F(2, 22) = 2.35$, *ns*; training: $F(1, 11) = 4.70$, *ns*) indicating that RTs increased with the magnitude of the rotation. None of the interactions were significant.

Insert Fig. 5 and Table 2 near here

ERs analysis. Figure 6 displays the average error rates (ERs) as a function of rotation between the stimuli to be matched for the two sessions. Table 2 presents the results of the linear regression analyses of ERs as a function of rotation for the two sessions.

Insert Fig. 6 near here

A three-way within-subject ANOVA including the factors of training, rotation and display mode was carried out on the ERs (means values collapsed across sessions.) The analysis showed significant main effects of rotation, *F* (4, 44) = 70.63, *p* < .001, and display mode, *F* (2, 22) = 21.91, *p* < .001, but no main effect of training, *F* (1, 11) < 1. ERs increased with rotation, as illustrated by the slopes in Table 2. ERs also varied as a function of display mode. Indeed, ERs were lower in the 3D than the 2D, *F* (1, 11) = 47.05, *p* < .001, or reversed 3D conditions, *F* (1, 11) = 16.94, *p* < .01. ERs were also marginally lower for the reversed 3D than the 2D condition, *F* (1, 11) = 4.71, *p* = .05.

The two-way interaction of rotation x display mode was significant, *F* (8, 88) = 4.28, *p* < .001. The contrast between the 2D and 3D conditions showed a significant interaction of rotation x display mode, *F* (4, 44) = 4.80, *p* < .01, which indicates that the rotation costs were weaker in the 3D than the 2D condition (see Table 2 and Fig. 6). The contrast between the 3D and reversed 3D conditions also showed a significant interaction of display mode x rotation, *F* (4, 44) = 3.60, *p* < .05. Rotation costs were weaker in the 3D than in the reversed 3D condition, as can be seen in Table 2 and Fig. 6. Finally, the contrast between the 2D and reversed 3D conditions also revealed a significant interaction of display mode x rotation, *F* (4, 44) = 4.11, *p* < .01, with weaker rotation costs in the reversed 3D condition than in the 2D condition (Table 2).

Discussion

Exp. 2 showed strong rotation effects on RTs and ERs. The rotation cost on ERs was weaker in the 3D display mode than in either the 2D or reversed 3D modes. Rotation costs were also slightly weaker in the reversed 3D than in the 2D condition. Training did not affect these interactions.

In Exp. 2, the effect of display mode on rotation costs remained constant across sessions 1 and 2. This contrasts with Exp. 1, in which performance was better with 3D displays in the first session, to then become better (with weaker rotation costs) with 2D and reversed 3D displays than with 3D display. In further contrast to Exp. 1, which showed no difference between the reversed 3D and 2D conditions, Exp. 2 showed better performance and slightly weaker rotation costs in the reversed 3D than the 2D display modes. The only difference between the two experiments was that the stimuli remained the same during both sessions in Exp. 1, while they were changed to a new set for the second session in Exp. 2. This methodological difference might account for the different patterns of results for the two experiments. However it is also possible that this difference was caused by differences between the subjects. Indeed, the data showed that the two groups were different since subjects of Exp. 2 made slightly more errors (17% vs 11%) than those of Exp. 1.

In the discussion of Exp. 1, we proposed that the reversal in the effect of normal stereopsis between sessions 1 and 2 was possibly due to changes produced by the development of perceptual expertise. Specifically, participants may have started the experiment by attempting to construct a global 3D representation of the shapes using stereo. However, stereo information was only valid in one third of the trials (i.e. in the 3D condition) whereas it was not in the remainder – stereo conflicted with the monocular depth cues in the 2D and reversed 3D conditions. We argue that this inconsistency between stereo and monocular depth cues eventually led participants to focus on depth-resistant 2D cues that they may have noted through exposure. That focus on learned 2D cues was prevented in Exp. 2 because a new set of stimuli was used in session 2. We argue that the main reason why a significant advantage of normal 3D was maintained throughout Exp. 2 is that 2D cues learned through expertise were unavailable in the second session of the experiment. As for the better performance with reversed 3D than with 2D displays in Exp. 2, it may be explained by the fact that participants may have learned to use the information provided by disparity magnitude while ignoring disparity sign. It remains however, that reversed 3D is a less efficient depth cue than normal 3D, as demonstrated by the present results.

The results from session one for both Exps 1 and 2 showed that normal stereo offers a performance advantage compared to reversed stereo or 2D. It may be noted that in Exp. 1, the effect of display mode occurred on RTs, while in Exp. 2, it occurred on ERs. Since the first session was the same for both experiments (except for the fact that a different set of stimuli was used for half the participants in Exp. 2), we might have expected the same results in both. However, we may also note that ERs were slightly higher in Exp. 2 (17.0%) than in Exp. 1 (11.0%). These higher ERs may have reduced the sensitivity of correct RTs to the effect of display mode while increasing that of ERs. Even though the effect of display mode did not occur on the same dependent variable in the two experiments, the consistent conclusion that emerges from the first session of Exps. 1 and 2 is that normal stereo offers a significant performance advantage over reversed stereo and 2D displays. This is congruent with the previous observations of Burke (2005) and it contradicts the claims by Biederman (1987) and Pizlo and colleagues (Chan, Stevenson, Li & Pizlo, 2006; Pizlo, 2008; Pizlo, Li & Francis, 2005; Pizlo, Li, Steinman, 2008; Pizlo & Stevenson, 1999) according to which stereopsis does not contribute to shape constancy.

Experiment 3

As argued in the discussion of Exp. 1, the inconsistency in the stereoscopic depth information available across the trials was probably involved in the reversal of the effect of normal stereo from facilitatory in session 1 to interference in session 2. This hypothesis is assessed in Exp. 3, which replicates Exp. 1 except for the fact that display mode is a between-subjects factor. This way, participants are exposed consistently either to 2D, normal stereo, or reversed stereo displays.

Methods

Participants. Twenty-four participants (15 females and 7 males, 23 right-handers and 1 left-hander) aged between 18 and 37 years old divided in three equal groups (one group for each display mode) took part in the experiment. All were naive as to the purpose of the experiment, were neurologically intact, and had normal or corrected visual acuity. No particular selection was applied with respect to gender, manual dominance, or level of education.

Procedure. The methods for Exp. 3 were very similar to Exp. 1, except that variations in display mode were between-subjects. For each group the experiment comprised two sessions of 200 trials each. The 100 “same” combinations across the four objects of Exp.1 were used again. Each of these combinations was presented twice. This gives a total of 200 “same” trials for each display mode. There were also 200 “different” trials for each display mode which were designed the same way as in Exp.1. The overall set of trials for each participant was distributed randomly across the two sessions.

Results

RTs that were more than 2.5 standard deviations away from a participant's average for a given condition were eliminated, resulting in the exclusion of a total of 197 data points for the complete experiment (2.0% of trials). The error trials were also excluded from the RTs analyses (10.6% of trials). The data showed no speed-accuracy trade-off since the correlation between correct RTs and ERs was positive for the three groups (2D: $r = .78$, $p < .05$, 3D: $r = .63$, $p < .05$, reversed 3D: $r = .74$, $p < .05$).

RTs analysis. Figure 7 displays the average correct RTs for "same" trials as a function of rotation for the two sessions. Table 3 presents the outcome of the linear regression analyses of RTs as a function of rotation for the two sessions. A three-way mixed ANOVA including the factors of training (session 1 and session 2), rotation (0, 20, 40, 60 and 80 degrees of rotation) and display mode (2D, stereo, reversed stereo) was carried out on correct RTs.

Insert Fig. 7 and Table 3 near here

Main effects of training, $F(1, 19) = 21.52$, $p < .001$, and rotation, $F(4, 76) = 63.82$, $p < .001$, were obtained. These indicated that RTs were shorter in the second session than in the first and that they increased with the degree of rotation. There was no significant main effect of display mode, $F(2, 19) = 1.23$, *ns*. However, the two-way interaction of rotation x display mode, $F(8, 76) = 3.36$, $p < .01$, was significant. It indicates that, for the means RT values collapsed across sessions, rotation costs were smaller for the 3D than the 2D condition, *F* (4, 56) = 4.86, *p* <.01 (see regression slopes in Table 3). However, there was no rotation cost difference between the 3D and the reversed 3D conditions, *F* (4, 56) = 1.08, *ns*, or between the reversed 3D and the 2D conditions, *F* (4, 56) < 1.

The three-way interaction of training x rotation x display mode, *F* (8, 76) = 2.28, *p* < .05, was also significant. It indicates that the impact of display mode on rotation costs was modulated by training. For the first session, only the rotation effect was significant, *F* (4, 80) = 36.58, *p* < .001. For the second session, there was a significant effect of rotation, *F* (4, 80) = 28.83, *p* < .001, and the two-way interaction rotation x display mode was also significant, *F* (8, 80) = 3.45, *p* < .001. The contrast between the 2D and 3D conditions revealed a significant difference in rotation costs, *F* (4, 56) = 50.73, *p* < .001, which were weaker in the 3D than the 2D display condition. The contrast between the 3D and reversed 3D conditions showed no significant rotation cost difference between the two conditions, *F* (4, 56) = 2.00, *ns*. Finally, the contrast between the 2D and reversed 3D conditions revealed a marginally significant difference in rotation cost, *F* (4, 56) = 2.28, *p* = .07, which tended to be weaker in the reversed 3D than the 2D condition (see Fig. 7 and regression slopes in table 3)

ERs analysis. Figure 8 displays the average ERs as a function of the degree of rotation between stimulus 1 and stimulus 2 for the two sessions. Table 3 presents the results of the linear regression analyses of ERs as a function of rotation for the two sessions. A three-way mixed ANOVA including the factors of training (session 1 and session 2), rotation (0, 20, 40, 60 and 80 degrees of rotation) and display mode (2D, 3D, reversed 3D) was carried out on ERs. Main effects of training, *F* (1, 21) = 13.25, *p* < .01, rotation, *F* (4, 84) = 43.40, *p* < .001, and display mode were obtained, *F* (2, 21) = 4.10, *p* < .05. These indicated that the ERs were lower in the second session than in the first, that they increased with the degree of rotation, and that they varied as a function of display mode. Indeed, ERs were significantly lower in the 3D than the 2D, F (1, 14) = 7.91, $p < .05$, or the reversed 3D conditions, F (1, 14) = 5.23, $p < .05$, but they did not differ between the reversed 3D and the 2D conditions, F (1, 14) < 1. The two-way interaction of rotation x display mode was also significant, F (4, 84) = 2.85, $p < .01$. No other interaction reached significance.

Pairwise contrasts between display modes on the means ER values collapsed across sessions showed a significantly weaker rotation cost for the 3D than either the 2D, F (4, 56) = 4.91, $p < .01$, or reversed 3D conditions, F (4, 56) = 4.34, $p < .01$ (see Table 2 and Fig. 5). However, the rotation effect did not differ between the 2D and reversed 3D conditions, F (4, 56) < 1.

Insert Fig. 8 near here

Discussion

As for Exp. 1, Exp. 3 showed strong rotation and training effects on RTs and ERs. In the first part of the experiment, there was no RT difference between the three groups. However, in the second part of the experiment, rotation costs on RTs were weaker for the 3D group than for the 2D group, and for the reversed 3D than the 2D group, but the difference between the 3D and the reversed 3D groups did not reach significance. The rotation costs on ERs were always weaker for the 3D group than for the 2D and reversed 3D groups, regardless of training.

The weaker rotation costs in the 3D condition than with other display modes in Exp. 3 are congruent with the findings reported by Burke (2005). This indicates that stereopsis contributes to shape constancy, contrary to what is proposed by the theories of Biederman (1987) and of Pizlo and colleagues (Chan, Stevenson, Li & Pizlo, 2006; Pizlo, 2008; Pizlo, Li & Francis, 2005; Pizlo, Li, Steinman, 2008; Pizlo & Stevenson, 1999).

The consistent performance advantage conferred by stereo in Exp. 3 also supports the hypothesis that the inversion of the display mode effect between sessions 1 and 2 in Exp. 1 was due in part to the inconsistency of the binocular disparity information to which participants were exposed in Exp. 1.

In Exp. 3, participants in the 3D group were possibly able to complete the experiment by attempting to construct a global 3D representation of the shapes using all the information available, including stereopsis, and did not need to rely on 2D invariant cues the extraction of which may be altered by stereo. However, participants in the 2D and reversed 3D groups may have ended up by using depth-resistant 2D cues that they were able to identify through stimulus exposure, as seems to have been the case in Exp. 1. Given stereo depth cues that are consistent with monocular depth cues, building a global 3D representation of shapes may be more efficient than relying on rotation-resistant 2D cues, which would explain why rotation costs were smaller for the 3D than 2D and reversed 3D groups.

We also note that rotation costs on RTs were weaker in the reversed 3D than the 2D condition in the second session. This observation indicates that even if the depth information carried by reversed stereo is inconsistent with monocular depth cues, it may contribute to shape constancy. A similar finding was also reported in Exp. 2 whereas no advantage for reversed 3D was observed in Exp. 1. These results show that in contexts favorable to the use of stereoscopic information for shape perception, observers may learn to use the depth information offered by reversed stereo. This issue is addressed further in the General Discussion.

General Discussion

The first goal of the study was to evaluate the contribution of stereopsis in shape constancy using reversed stereopsis, a condition that had never been used before for this purpose. The second goal was to understand how the development of perceptual expertise can affect the contribution of stereopsis. In the three experiments, we used a sequential matching task with bent paperclips rotated in depth that were presented without stereopsis, or with normal or reversed stereopsis. In Exp. 1, the variations of the display mode were within-subject. In the first session, performance was better for the 3D display mode than for the 2D or reversed 3D display modes. However, in the second session, the effect of display mode reversed and rotation costs were larger in the 3D display mode than in the other conditions. Exp. 2 used the same methods as Exp. 1 except that there was a switch of stimuli between the first and the second sessions. The rotation costs were weaker in the 3D display mode than with reversed 3D or 2D displays, and slightly weaker with reversed 3D than with 2D displays. In Exp. 3, the factor of display mode was between-subjects. The rotation costs were weaker in the 3D display mode than with reversed 3D or 2D displays for the entire experiment, and they were slightly weaker with reversed 3D than in the 2D condition.

The role of stereopsis in shape constancy

The present study showed that stereopsis may help to reduce rotation costs in the discrimination of shapes that do not meet the conditions for rotation invariance proposed Biederman and Gerhardstein (1993) or the simplicity constraints proposed by Pizlo and Stevenson (1999). This finding is consistent with the previous results of Bennett and Vuong (2006), Burke (2005), and Burke et al. (2007). It may also be conceived as congruent with Biederman's theory (Biederman, 1987; Biederman & Gerharstein, 1993) since the present stimuli do not meet the conditions for orientation invariance.

While our results demonstrate that stereoscopic information may contribute to shape constancy, they also show that stereo may sometimes bring a performance cost in shape discrimination, as observed in session 2 of Exp. 1. As argued previously, the development of perceptual expertise may lead observers to focus on the most useful sources of information for the task at hand (Palmeri et al., 2004). In Exp. 1, the information provided by binocular disparity was inconsistent and this seems to have led participants to seek other sources of information upon which to base their responses. With sufficient expertise with the stimuli in session 2 of the experiment, observers apparently rejected binocular disparity information to focus on more informative sources of information, which we assume are 2D rotation-resistant cues. The joint contribution of the variation of the consistency of binocular disparity information across the trials and object-specific expertise in producing this phenomenon is demonstrated in Exps. 2 and 3. Thus, in Exp. 2, there was variation of the consistency of binocular disparity information across the trials as in Exp. 1 but the development of object-specific expertise was short-circuited by changing stimulus sets between sessions 1 and 2. The results of Exp. 2 showed a consistent facilitatory contribution of normal stereo across sessions 1 and 2. The same outcome followed in Exp. 3, in which the development of object-specific expertise was possible, but with no variation of the consistency of the information value of binocular disparity for observers since display mode was manipulated as a between-subjects factor.

One may nevertheless wonder why normal 3D became a source of interference in session 2 of Exp. 1 instead of just having no effect. In this regard, it is worth noting that Pasqualotto and Hayward (2009) have shown before that particular test conditions may lead to a stereo disadvantage in the recognition of familiar objects. In the particular context of their experiment, they proposed that participants may have relied on the 2D outline shape to perform the recognition task required, and that the extra specificity provided by stereoscopic disparity seems to have altered the extraction of the 2D information. It is possible that normal stereo may have caused the same problem for the extraction of the 2D rotation-invariant cues in Exp. 1, which would explain the longer RTs and greater ERs in the 3D condition in the second session of that experiment.

The impact of reversed stereopsis

The impact of reversed stereopsis on shape constancy has never been studied before. The results of Exp. 1 show that when stereoscopic information is inconsistent with other depth cues, as was the case in the reversed 3D condition, it is probably ignored in the discrimination of rotated shapes. Indeed, had reversed stereo not been ignored, the results in the 2D and reversed 3D conditions should have differed in some way, which they did not. However, reversed stereo was not ignored in Exps. 2 and 3 since it brought a performance advantage relative to the 2D display mode, and it even brought an equal advantage than normal stereo in the RT of Exp 3.

The rejection of reversed stereopsis in Exp. 1 is congruent with the impact of inconsistent depth cues in surface evaluation tasks. Indeed, when stereopsis is inconsistent with the depth cues carried by motion, texture or edges, these more useful sources of information are retained and stereo is rejected (Norman & Todd, 1995; Saunders & Backus, 2006, and Stevens, Lees & Brookes, 1991). Our results might also seem consistent with those of Ichikawa and Egusa (1993), who showed that long term adaptation to reversed stereopsis leads to a change in the weight of the different depth cues. Specifically, after adaptation, reversed stereopsis was ignored and only monocular depth cues were used to judge depth.

Conversely, the benefits of reversed 3D displays in Exps. 2 and 3 are congruent with the findings of Shimojo and Nakajama (1981), who found that with the long-term wearing of depth reversal spectacles, participants became able to evaluate accurately the depth of pyramid shapes. Their participant wore the spectacles for ten days but the effects of the adaptation appeared at the beginning of the third day Furthermore, when the glasses were removed, they perceived the same shapes in reversed depth. We suggest that a phenomenon of the same kind may have occurred in Exps. 2 and 3. However, it was long-term adaptation in Shimojo and Nakajama, and the possible adaptation we found here was short-term adaptation and occurred in less than 45 minutes rather than in two days. No other study seems to have shown yet a rapid adaptation to reversed binocular disparity.

It is interesting to note that even though reversed stereo is inconsistent with monocular depth cues it brings relevant information. Specifically, its information value is the same as that offered by normal stereo, except that it is reversed in sign. Once participants are adapted to this sign reversal, reversed stereo may become just as informative as regular stereo. This is consistent with the results of Matthews, Hill and Palmisano (2011), who showed that for the evaluation of face depth, participant could use the disparity magnitude without using the disparity sign. However to the best of our knowledge, there is no physiological model of disparity in which magnitude and sign are encoded separately. This remains to be explored.

The independant use of the disparity magnitude and sign is however not the only possible explanation of the advantage of reversed stereo over 2D. One thing to take in consideration about the stimuli used in these experiments is that if we remove all the monocular information, except their silhouette, the shapes that correspond to the stimuli viewed with reversed disparity are the mirror reflection of the stimuli viewed with normal disparity, but that rotated in the opposite direction. It is possible that the monocular information weight was reduce or that it was ignored in Exps 2 and 3, thus reducing or eliminating the conflict between monocular and binocular cues and thus conferring an advantage of reversed 3D over 2D. In Exp 1, since participants were probably focusing on 2D cues, monocular information was not ignored and there reversed stereoscopic information probably was.

In a nutshell, these results show that reversed stereopsis may contribute to shape constancy under particular conditions. However, the combined results of the three experiments indicate that its use depends of the constraints of the task, and more research would be necessary to characterize correctly the impact of reversed stereopsis in shape perception.

Methodological considerations

The results of the present study show that it is important to be careful when comparing the effects of normal stereopsis with absent or reversed stereopsis. In Exp. 1, the random distribution of display modes across trials led observers to focus on sources of information whose processing was altered by normal stereo. Under a superficial analysis of the results of Exp. 1 (e.g. without consideration of the training factor) and in the absence of our other experiments, we would have been led to conclude erroneously that stereoscopic information is irrelevant to shape constancy. If the different display conditions had been blocked instead of being randomly distributed, this might not have occurred. However, a blocked design may not be the ideal solution given that the effects of practice and familiarity with the stimuli would contaminate the factor of display mode.

Another aspect of our methods which may have favored the development of object-specific expertise in Exp. 1 is the long duration of the stimuli (2,500 ms for the first stimulus in the sequence, and up to 5,000 ms for the second). This may have allowed participants to focus on the details of the shapes and thus discover rotation-resistant 2D cues, upon which their strategy was apparently based in the second session of Exp. 1. It is possible that the pattern of results would have been different with a shorter stimulus exposure duration.

Further studies

It is well known that the ventral pathway processes visual objects, forms, texture and color, and that the dorsal pathway is implicated in spatial localisation and in action guidance (Goodale & Milner, 1992; Goodale et Westwood 2004; Ungerleider & Mishkin, 1982). However it has been found that some neurons in the ventral pathway show some sensitivity to binocular disparity (Hinkle & Connor, 2001; Uka, Tanaka, Yoshiyama, Kato & Fujita, 2000), and that V5, part of the dorsal pathway, is sensitive to 3D shapes created by the combination of different depth cues (Welchman, Deubelius, Conrad, Bülthoff and Kourtzi, 2005). These results are consistent with the contribution of stereopsis to shape constancy demonstrated here since both pathways seem to integrate the depth cues to achieve shape representation. We already know that the recognition of rotated objects is associated with an increase in the activity of some ventral areas as a function of the size of the rotation (Gauthier et al, 2002) but it would be interesting to evaluate how stereo may affect that activity.

Conclusion

In summary, the present findings suggest that the depth information provided by stereo can be integrated to shape representations to sustain shape constancy. This, however, requires stereoscopic information validity to be minimally consistent across the trials. Otherwise, the development of perceptual expertise leads to a focus on other sources of information for which stereo can be deleterious.

Acknowledgements

The research reported in this article was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (grant number 203613-2008) to Martin Arguin.

References

Arguin, M. & Saumier, D. (2004) Independent processing of parts and of their spatial organization in complex visual objects, *Psychological Science, 15*, 629-633.

Bennett, D. J., & Vuong, Q. C. (2006) A stereo advantage in generalizing over changes in viewpoint on object recognition tasks, *Perception and Psychophysics, 68*, 1082-1093.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding, *Psychological Review, 94*, 115-147.

Biederman I, et Gerharstein, P.C. (1993) Recognizing depth-rotated objects: Evidence and conditions for three-dimensional view-point invariance, *Journal of experimental Psychology: Human perception and performance, 19*, 1162-1182.

Biederman, I., & Ju, G. (1988) Surface versus edge-base determinants of visual recognition, *Cognitive Psychology, 20*, 38-64.

Burke, D. (2005) Combining disparate views of objects: Viewpoint costs are reduced by stereopsis, *Visual Cognition, 12*, 201-213.

Burke, D., Taubert, J., & Higman, T. (2007) Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces, *Vision Research, 47*, 2164-2169.

Chan, M. W., Stevenson, A. K., Li, Y., & Pizlo, Z. (2006) Binocular shape constancy from novel views: The role of a priori constraints, *Perception & Psychophysics, 68*, 1124-1139.

Edelman S., & Bülthoff H. H. (1992) Orientation dependence in the recognition of familiar and novel views of three-dimensional objects, *Vision Research, 32*, 2385-2400.

Gauthier , I., Hayward, W.J., Tarr, M.J., Anderson, A.W., Skudlarski, P. & Gore, J.C. (2002) BOLD activity during mental rotation and viewpoint-dependant object recognition, *Neuron, 34*, 161-171.

Goodale, M.A. & Milner, A.D. (1992) Separate visual pathways for perception and action, *Trends in Neuroscience*, 15, 20-25.

Goodale, M.A. & Westwood, D.A.(2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action, Current Opinion in Neurobiology, 14, 203-211.

Hayward, W. G., & Tarr, M. J. (1997) Testing Conditions for Viewpoint Invariance in Object Recognition, *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511-1521.

Hinkle, D.A. & Connor, C.E. (2001) Disparity tuning in macaque area V4, *Neuroreport, 12*, 365-369.

Ichikawa, M. & Egusa, H. (1993) How is depth perception affected by long-term wearing of left-right reversing spactacles?, *Perception, 22*, 971-984.

Ichikawa, M., Eguase, H., Nakatsuka, M., Amano, J., Ueda, T., Tashiro, T. (2003), Modification of depth and distance perception caused by long-term wearing of left right reversing spectacles, *Perception*, *32*.

Marr, D. (1982). Vision. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978) Representation and recognition of three dimensional shapes, *Proceeding of the Royal Society of London, B, 200*, 269-291.

Matthews, H. Hill, H., Palmisano, S. (2011) Binocular disparity magnitude affects perceived depth magnitude despite inversion of depth order, *Perception, 40*, 975-988

Norman, J. F. & Todd, T. T. (1995) The perception of 3-D structure from contradictory optical patterns, *Perception & Psychophysics, 57*, 826-834.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004) Computational approaches to the development of perceptual expertise, *Trends in Cognitive Neuroscience, 8*, 378-386.

Pasqualotto, A. & Hayward, W. G. (2009) A stereo disadvantage for recognizing rotated familiar object, *Psychonomic Bulletin & Review*, *16*, 832-838.

Peissig, J. J., Wasserman, E. A., Young, M. E., & Biederman, I. (2002). Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons. *Vision Research, 42*, 2051–2062.

Pentland, A. (1985). On describing complex surface shapes. *Image and Vision Computing, 3*, 153–162.

Pizlo, Z., Li, Y., & Francis, G. (2005) A new look at binocular stereopsis, *Vision Research, 45*, 2244-2255.

Pizlo, Z., Li, Y., & Steinman, R. M. (2008) Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose, *Spatial Vision, 21*, 495-508.

Pizlo, Z. & Stevenson, A. (1999) Shape constancy from novel views, *Perception & Psychophysics, 61*, 1299-1307.

Saunders, J. A. & Backus, B. T. (2006) Perception of surface slant from oriented textures, *Journal of Vision, 6*, 882-897.

Shimojo, S., & Nakajama, Y. (1981) Adaptation to the reversal of binocular depth cues: Effects of wearing left-right reversing spectacles on stereoscopic depth perception, *Perception, 10*, 391-402.

Stevens, K. A., Lees, M., & Brookes, A. (1991) Combining binocular and monocular features, *Perception, 20*, 425-440.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geons structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1494–1505.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8*, 282 289.

Tarr, M. J., & Pinker, S. (1989) Mental rotation and orientation-dependence in shape recognition, *Cognitive Psychology, 21*, 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1*, 253–256.

Tarr, M. J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape recognition: Further issues, *Psychological Science, 2*, 207–209.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent, *Nature Neuroscience, 1*, 275-277.

Uka, T., Tanaka, H., Yoshiyama, K., Kato, M. & Fujita, I. (2000) Disparity selectivity of neurons in monkey inferior temporal cortex, *Journal of Neurophysiology, 84*, 120-132.

Ungerleider, L.G., & Mishkin, M. (1982) Two cortical systems, *Analysis of Visual Behavior,* Cambridge, MA, MIT Press, 549-586.

Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H. & Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortx, *Nature Neuroscience, 8(6),* 820-827.

Yellott, J. I., & Kaiwi, J. L. (1979) Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective, *Perception, 8*, 135-142.

Footnotes

[1] These conditions state that the objects must possess a geon structural description, these descriptions must remain the same across the different viewpoint and they must differ across objects.

TABLES

Table 1

Linear regression of "same" trials RTs and ERs as a function of rotation for each session and each display mode in Exp. 1

		Response times			Error rates		
Session	Condition	Slope (ms/deg)	Intercept (ms)	R-square	Slope (p/deg)	Intercept (p)	R-square
1	2D	4.51	623	.96	.0042	.00	.97
	3D	3.14	485	.96	.0030	.01	.93
	Rev. 3D	4.14	610	.96	.0024	.02	.94
2	2D	3.47	555	.97	.0027	-.01	.82
	3D	4.36	546	.98	.0038	-.02	.94
	Rev. 3D	3.36	552	.93	.0026	-.01	.97
Mean	2D	3.99	589	.97	.0034	-.01	.96
	3D	3.75	516	.98	.0034	.00	.96
	Rev. 3D	3.75	581	.96	.0025	.01	.99

Table 2

Linear regression of "same" trials RTs and ERs as a function of rotation for each session and each display mode in Exp. 2

		Response times			Error rates		
Session	Condition	Slope (ms/deg)	Intercept (ms)	R-square	Slope (p/deg)	Intercept (p)	R-square
1	2D	3.24	780	.96	.0059	.07	.98
	3D	3.85	733	.98	.0040	.03	.96
	Rev. 3D	2.77	773	.90	.0057	.04	.99
2	2D	2.71	674	.83	.0070	-.01	.98
	3D	2.70	645	.99	.0045	.00	.98
	Rev. 3D	3.27	673	.93	.0055	.02	.95
Mean	2D	2.97	727	.91	.0064	.03	.99
	3D	3.28	689	.99	.0042	.01	.97
	Rev. 3D	3.02	723	.99	.0056	.03	.99

Table 3

Linear regression of "same" trials RTs and ERs as a function of rotation for each session and each display mode in Exp. 3.

		Response times			Error rates		
Session	Condition	Slope (ms/deg)	Intercept (ms)	R-square	Slope (p/deg)	Intercept (p)	R-square
1	2D	4.18	755	.97	.0058	.01	.99
	3D	3.25	680	.94	.0014	.04	.89
	Rev. 3D	4.07	708	.93	.0047	-.01	.89
2	2D	5.13	611	.99	.0035	-.01	.98
	3D	1.81	617	.59	.0021	.00	.83
	Rev. 3D	2.80	658	.96	.0034	-.02	.88
Mean	2D	4.65	683	.96	.0046	.00	.99
	3D	2.53	648	.84	.0018	.02	.95
	Rev. 3D	3.43	683	.95	.0041	-.01	.93

FIGURE CAPTIONS

FIG.1. Objects used as stimuli in the three experiments

FIG. 2. Average correct RTs in Exp. 1 as a function of rotation. A) Session 1. B) Session 2. C) Mean of sessions 1 and 2.

FIG. 3. Average ERs in Exp. 1 as a function of rotation. A) Session 1. B) Session 2.C) Mean of sessions 1 and 2.

FIG. 4. New set of stimuli used in Exp. 2.

FIG. 5. Average correct RTs in Exp. 2 as a function of rotation. A) Session 1. B) Session 2.C) Mean of sessions 1 and 2.

FIG. 6. Average ERs in Exp. 2 as a function of rotation. A) Session 1. B) Session 2.C) Mean of sessions 1 and 2.

FIG. 7. Average correct RTs in Exp. 3 as a function of rotation. A) Session 1. B) Session 2.C) Mean of sessions 1 and 2.

FIG. 8. Average ERs in Exp. 2 as a function of rotation. A) Session 1. B) Session 2.C) Mean of sessions 1 and 2.

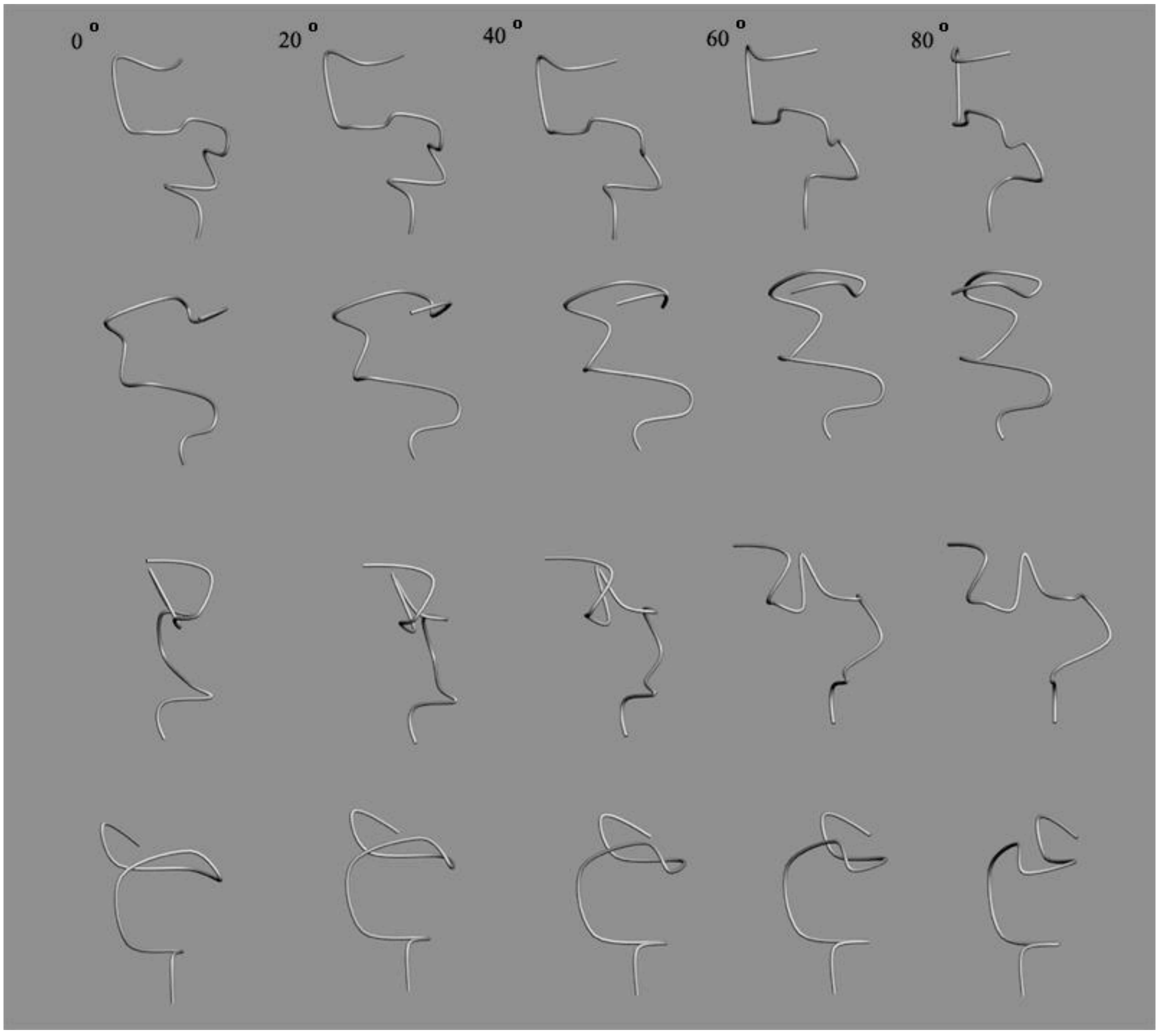


Fig. 1

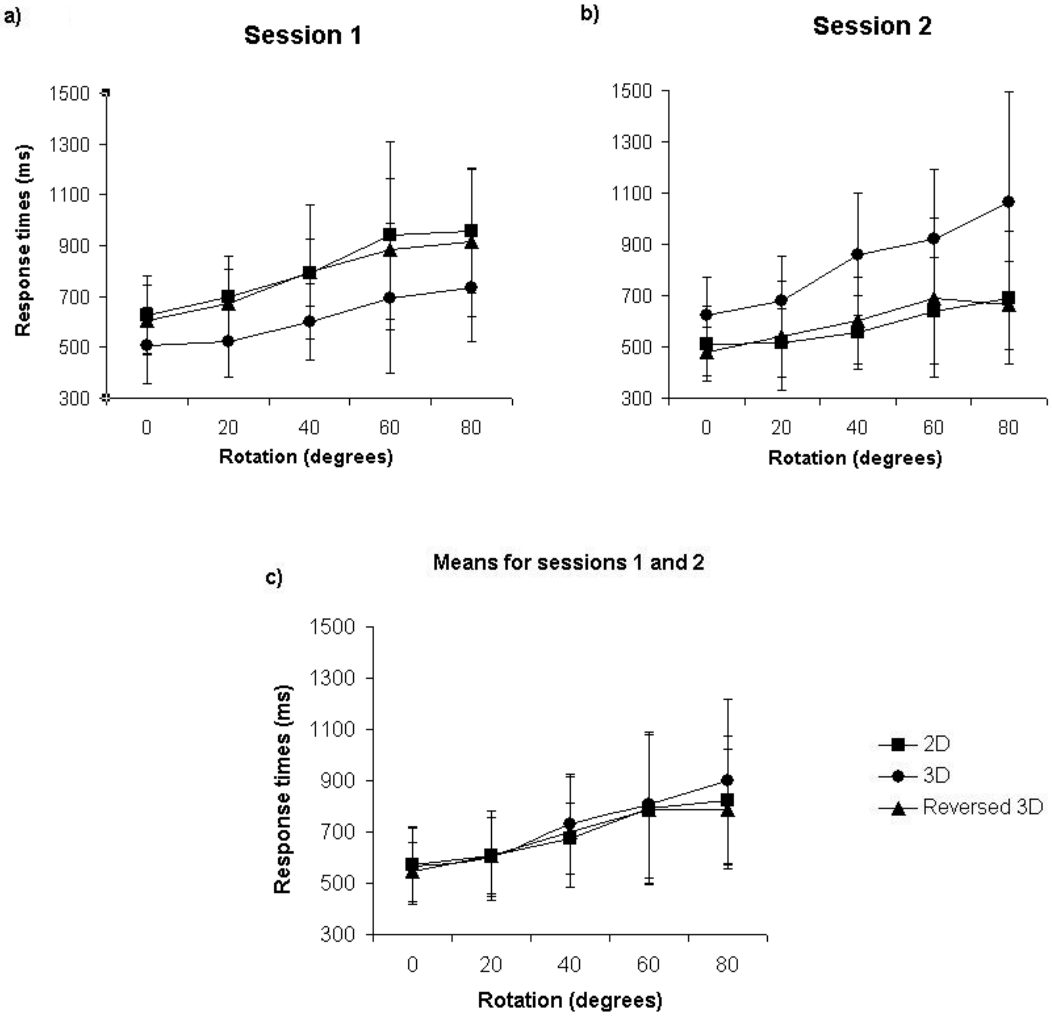

a)
Session 1
b)
Session 2
c)
Means for sessions 1 and 2
Response times (ms)
1500
1300
1100
900
700
500
300
Rotation (degrees)
0
20
40
60
80
2D
3D
Reversed 3D

Fig. 2

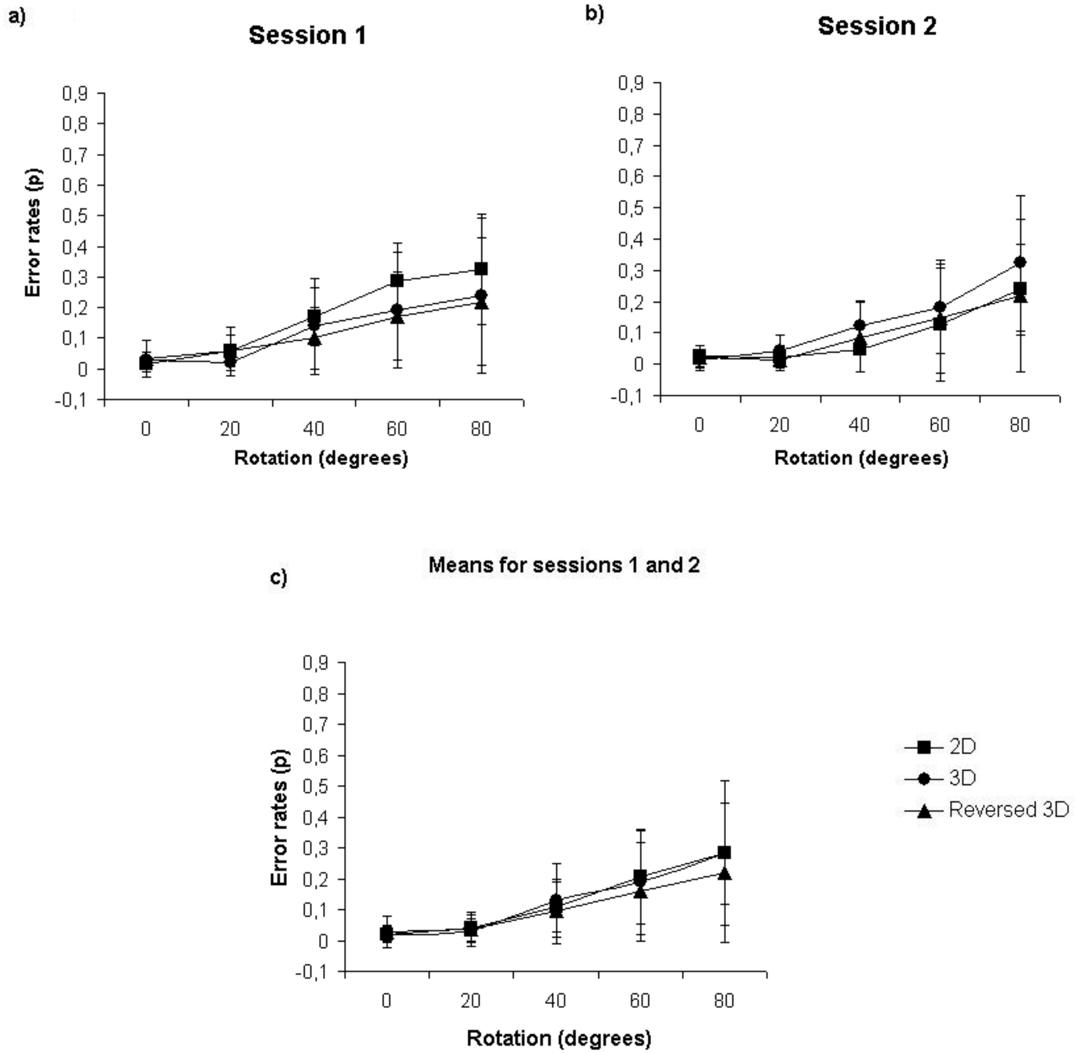

a)
Session 1
b)
Session 2
c)
Means for sessions 1 and 2
Error rates (p)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
2D
3D
Reversed 3D

Fig. 3

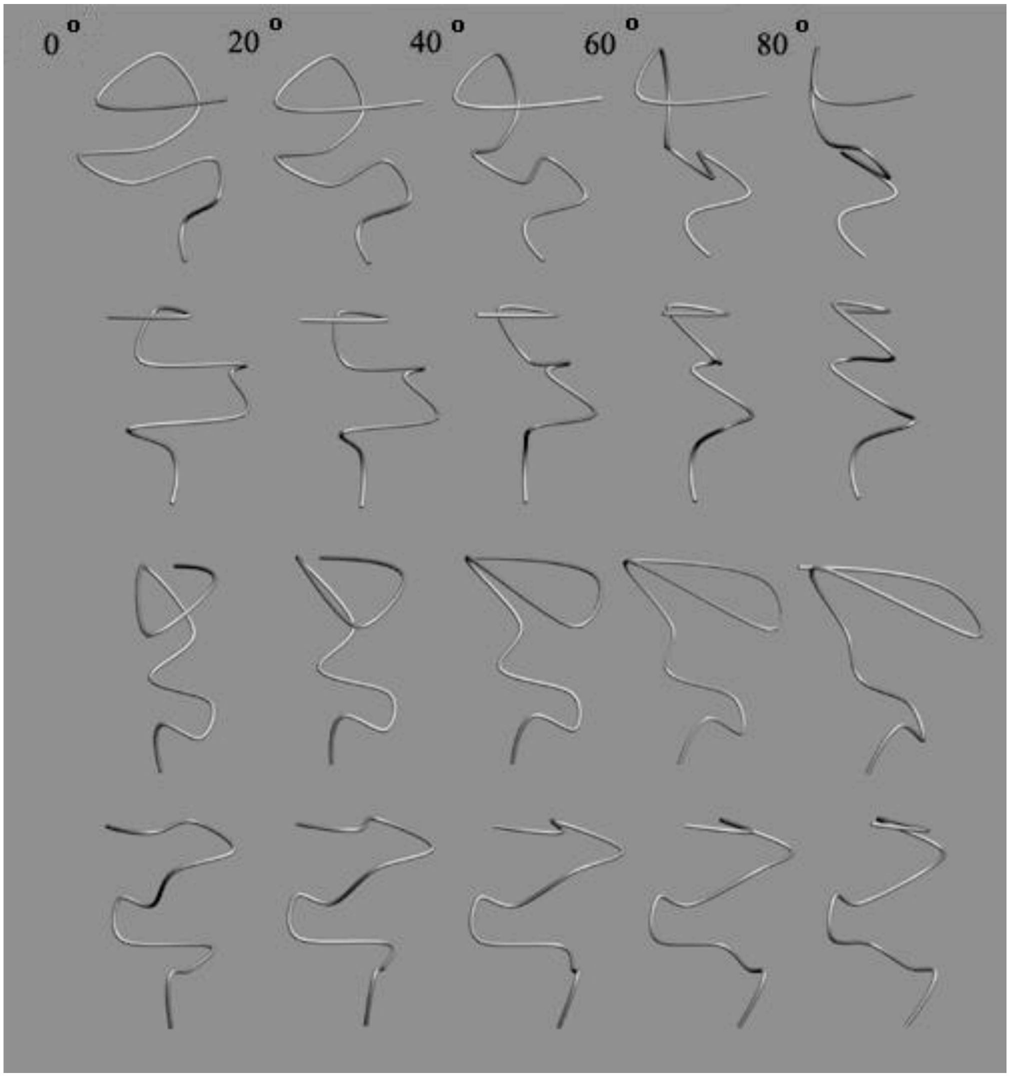


Fig. 4

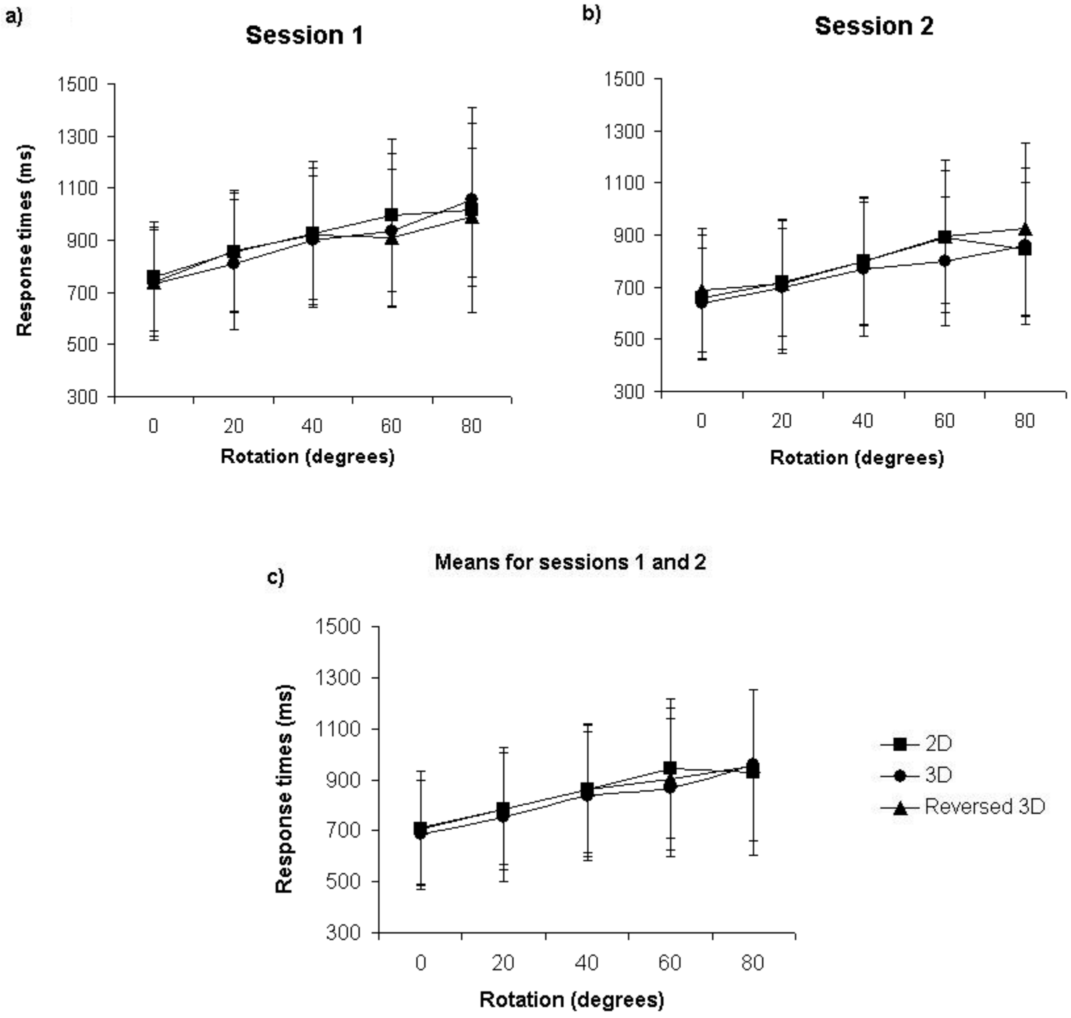

a)
Session 1
b)
Session 2
c)
Means for sessions 1 and 2
Response times (ms)
Rotation (degrees)
1500
1300
1100
900
700
500
300
0
20
40
60
80
2D
3D
Reversed 3D

Fig. 5

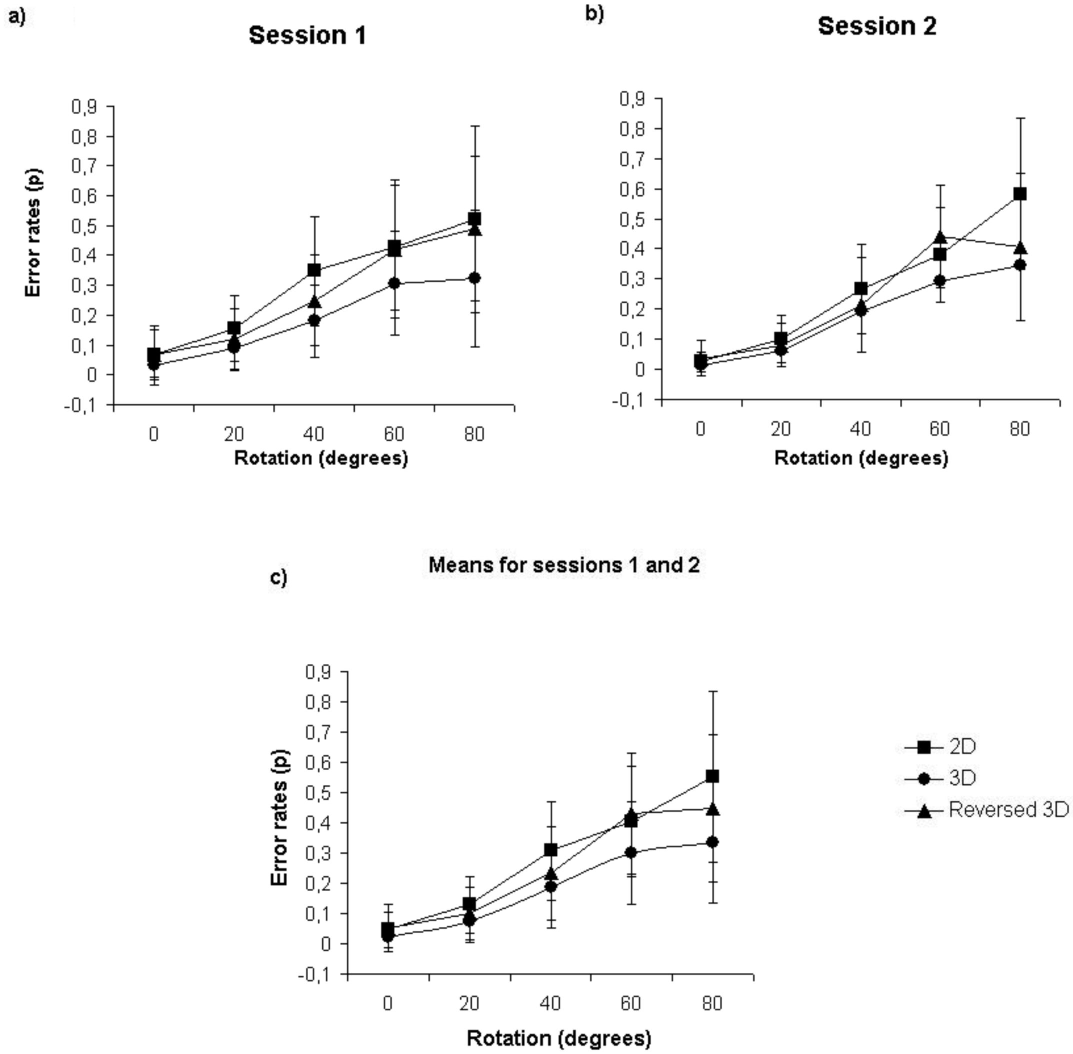

a)
Session 1
Error rates (p)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
b)
Session 2
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
c)
Means for sessions 1 and 2
Error rates (p)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
2D
3D
Reversed 3D

Fig. 6

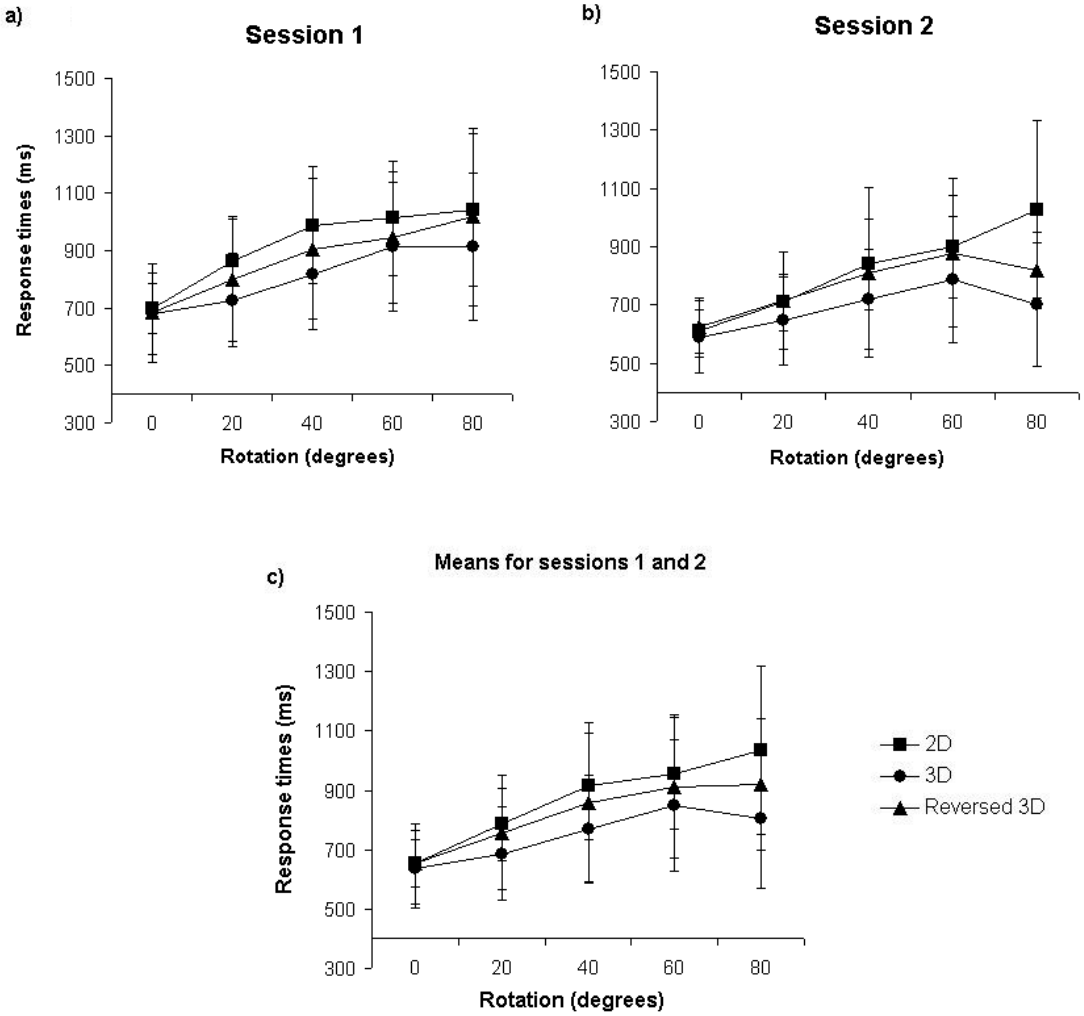

a)
Session 1
b)
Session 2
c)
Means for sessions 1 and 2
Response times (ms)
1500
1300
1100
900
700
500
300
0
20
40
60
80
Rotation (degrees)
2D
3D
Reversed 3D

Fig. 7

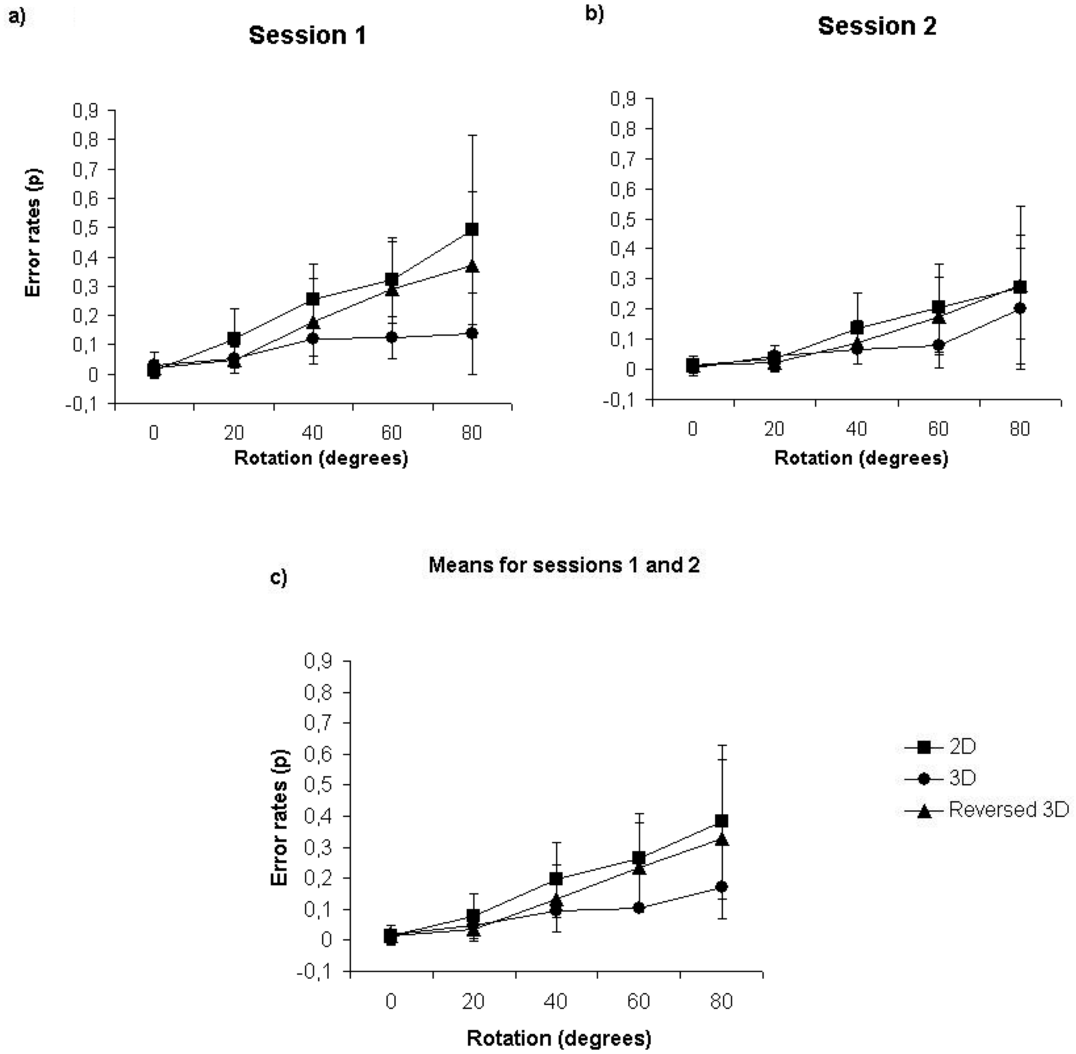

a)
Session 1
Error rates (p)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
b)
Session 2
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
c)
Means for sessions 1 and 2
Error rates (p)
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
0
20
40
60
80
Rotation (degrees)
2D
3D
Reversed 3D

Fig. 8

Discussion générale

Résumé des buts et des résultats

Article 1. Les buts du premier article étaient d'évaluer l'impact de la stéréoscopie dans la reconnaissance de forme et de déterminer si la disparité binoculaire inversée est une condition intéressante pour mesurer le rôle de la stéréoscopie dans la perception des formes. L'expérience consistait en une tâche de reconnaissance de forme (i.e. indiquer si des formes étaient nouvelles ou familières). Les formes pouvaient être présentées avec disparité binoculaire, sans disparité entre les vues présentées aux deux yeux ou avec disparité binoculaire inversée. La variation du mode de présentation était inter-sujet. Les résultats ont montré que les performances étaient moins bonnes dans la condition de stéréoscopie inversée que dans les conditions de stéréoscopie normale ou d'absence de disparité entre les vues présentées aux deux yeux. Cela indique que la stéréoscopie est utilisée par les processus sous-tendant la reconnaissance de forme.

Le fait que la stéréoscopie soit prise en compte lors de la reconnaissance de forme est incongruent avec les théories de Biederman (1987) et de Pizlo (2008). En effet, les formes utilisées répondaient à leurs différentes contraintes, et les contours à eux seuls étaient suffisants pour bien réussir la tâche. En ce sens, les données concordent avec les résultats de Bennett et Vuong (2006), Burke (2005) et Burke et al. (2007) qui ont montré que la stéréoscopie contribuait à la constance de forme. De plus, bien que Julesz (1960, 1971) ait montré qu'il était possible de reconnaître des formes uniquement définies par la disparité binoculaire, la contribution de la stéréoscopie lorsque l'information de profondeur monoculaire est riche n'était pas encore claire. Ces résultats appuient l'idée que la stéréoscopie contribue à la reconnaissance de forme malgré la richesse de l'information monoculaire.

La stéréoscopie inversée a permis de mesurer un effet qui n'aurait pas été détecté si seules les conditions de disparité normale et d'absence de disparité avaient été comparées. Elle amène donc une nouvelle sensibilité qui pourrait permettre de mieux comprendre le rôle de la stéréoscopie et comment elle est utilisée par les processus sous-tendant la représentation des formes.

Toutefois, on pourrait interpréter les données autrement et proposer que l'absence de différence entre les conditions 2D et 3D ne soit pas liée à un effet plancher, mais plutôt au fait que la stéréo n'amène aucun avantage. Dans ce cas, l'effet délétère de la 3D inversée indiquerait que la stéréo est prise en compte uniquement lorsqu'elle est très inconsistante avec les informations monoculaires. Cependant, les présents résultats ne permettent pas de départager les deux hypothèses. Il serait toutefois intéressant de refaire la tâche en augmentant le niveau de difficulté (en diminuant le contraste des stimuli par exemple) afin d'augmenter les taux d'erreur et de vérifier si l'absence de différence entre les conditions 2D et 3D était liée à un effet plancher ou non.

Article 2. Le but de l'article 2 était d'évaluer l'intégration de l'ombrage et de la stéréoscopie dans la perception du relief. Une tâche de jugement du relief a été effectuée. Les participants devaient juger si des formes au relief ambigu étaient convexes ou concaves. L'illumination des formes pouvait venir du haut à gauche, du haut à droite, du bas à gauche et du bas à droite, et les stimuli pouvaient être présentés avec disparité binoculaire, avec disparité binoculaire inversée ou sans disparité entre les vues présentées aux deux yeux. Les résultats ont montré que les participants répondaient « convexe » plus souvent lorsque les formes étaient présentées avec stéréoscopie que sans stéréoscopie, et ont répondu que les formes étaient convexes plus souvent lorsqu'elles étaient présentées en 2D qu'avec stéréoscopie inversée. De plus, le taux de réponse «convexe » était plus haut lorsque les formes étaient éclairées par le haut que par le bas. Le fait que l'illumination vienne de la gauche ou de la droite n'a rien changé aux réponses. Finalement, il n'y a pas eu d'interaction entre le mode de présentation et la direction d'illumination dans les réponses. Ces résultats indiquent que la stéréoscopie et l'ombrage contribuent indépendamment à la représentation du relief des formes.

Les résultats de l'article sont congruents avec l'effet de l'a priori d'illumination par le haut tel que montré par Adams (2007), Connor (2001), Gerardin, Montalembert et Mamassian (2007) et Kleffner et Ramachandran (1992). Par contre, le fait que la lumière vienne de la gauche ou de la droite n'a pas influencé les résultats. Ceci appuie les données de l'Exp. 1 de McManus, Buckman et Wolley (2004), qui ont utilisé des stimuli présentés pour une durée indéfinie. Par contre, nos résultats sont incongruents avec l'hypothèse d'un biais pour la lumière provenant de la gauche qui était suggéré par les études ayant fait usage de courtes durées de présentation (Mamassian & Goutcher, 2001; McManus et al. – Exp. 2; Sun & Perona, 1998). Puisque nos stimuli demeuraient visibles jusqu'à la réponse des participants, ceci suggère que l'effet de la direction d'illumination gauche/droite est liée à la durée de présentation. Il est possible qu'avec de plus longues durées d'exposition, les participants adoptent une stratégie différente puisqu'ils ont plus de temps pour analyser l'information. À cet égard, Sun et Perona ont montré que le biais d'illumination vers la gauche était corrélé à la dominance manuelle. Ceci suggère la possibilité que la latéralisation hémisphérique soit impliquée dans ce biais et que de longues durées de présentation laissent plus de temps pour un transfert inter-hémisphérique qui entraînerait sa disparition.

Les effets de direction d'illumination et de mode de présentation étaient additifs, ce qui signifie que l'ombrage et la stéréoscopie avaient une contribution indépendante aux processus impliqués dans le jugement des formes. Ces résultats vont à l'encontre de ceux obtenus par Braunstein (1986), Norman et Todd (1995), Saunders et Backus (2006), et Stevens et Brookes (1998), qui ont montré des interactions entre les différentes sources d'information de profondeur. Ainsi, en cas d'incongruence entre les indices, plus de poids est accordé à ceux qui sont les plus informatifs, ceux-ci pouvant même éliminer toute contribution des autres indices. Par contre, nos observations appuient le modèle de Landy et al. (1995), qui suppose que les différents indices de profondeur sont intégrés linéairement Celles-ci confirment également les données de Dosher et al. (1986), qui montrent que la stéréo et la covariance de luminance de proximité sont intégrées linéairement.

Article 3. Le but de l'article 3 était de vérifier l'impact de la stéréoscopie dans la constance de forme. Trois expériences d'appariement séquentiel de stimuli en forme de trombones tordus ayant subi des rotations et pouvant être présentés en 2D, 3D et 3D inversée ont été menées. Dans la première expérience, les variations du mode de présentation étaient intra-sujets. Dans la première moitié de l'expérience, les performances étaient meilleures lorsque les formes étaient présentées avec stéréoscopie que sans disparité entre les vues présentées aux deux yeux ou avec stéréoscopie inversée. Toutefois, c'était le patron inverse pour la seconde moitié de l'expérience où les coûts de rotation sont devenus plus grands pour la condition de stéréoscopie normale que pour les conditions de 2Det de stéréoscopie inversée. Il est proposé que ce renversement dans l'effet de la disparité binoculaire est causé par une adaptation stratégique liée au développement de l'expertise perceptuelle et à l'inconsistance de l'information de disparité binoculaire présentée aux participants. Il est postulé qu'au début de l'expérience, les participants tentaient de créer des représentations globale 3D des formes en faisant usage notamment de la stéréoscopie. Toutefois, puisque la stéréoscopie était incongruente avec l'information monoculaire dans le deux tiers des essais, les participants ont possiblement appris à détecter des détails 2D résistants à la rotation et à les utiliser pour reconnaître les formes. Une telle possibilité s'accorde avec la notion voulant que le développement de l'expertise perceptuelle fasse en sorte que l'attention est déplacée vers les caractéristiques les plus informatives des stimuli (Palmeri, Wong & Gauthier, 2004). Dans notre expérience, il est possible que la stéréoscopie ait affecté ou ralenti le traitement de ces détails 2D constituant une meilleure source d'information, d'où le désavantage qu'elle a entrainé dans la deuxième partie de l'expérience. De plus, puisqu'il n'y avait pas de différence entre les conditions 2D et 3D inversée, il semble que la stéréoscopie incongruente avec les indices de profondeur monoculaires ait été ignorée. Afin de vérifier la possibilité de cette adaptation stratégique liée à l'exposition aux différents modes de présentation, l'expérience 2 a utilisé la même tâche que l'expérience 1, mais avec un changement de stimuli entre les séances 1 et 2. Une telle manipulation empêche l'intervention d'une expertise spécifique aux stimuli à la séance 2 de l'expérience et les résultats démontrent un maintien de l'avantage pour la condition de disparité normale à travers toute la durée de l'expérience. On note également que les coûts de rotation étaient plus faibles pour la condition de stéréoscopie inversée qu'avec une absence de disparité entre les vues présentées aux deux yeux. L'expérience 3 a utilisé la même tâche que l'expérience 1, mais avec une variation inter-sujet du mode de présentation. Les résultats ont montré que les coûts de rotation sur les temps de réponse étaient plus faibles avec la stéréoscopie normale qu'avec l'absence de disparité, mais ne différaient pas d'avec la stéréo inversée. Aux niveaux des taux d'erreurs ils étaient plus faibles avec la stéréo normale qu'avec la stéréo inversée. De plus, ils étaient plus faibles avec la stéréo inversée que dans la condition 2D. Ces résultats indiquent que la stéréoscopie contribue à la constance de forme. De plus, la stéréoscopie inversée semble y contribuer aussi. Cette dernière est incongruente avec l'information monoculaire dans le sens ou l'information qu'elle offre est de signe opposé à l'information transmise par les indices monoculaires. Toutefois, de manière congruente aux données de Shimojo et Nakajama (1981), nos résultats indiquent qu'il est possible de s'adapter à l'inversion de ce signe.

L'avantage de la stéréoscopie appuie les observations précédentes de Bennett et Vuong (2006), Burke (2005), et Burke et al. (2007), qui avaient montré que la stéréoscopie peut réduire les coûts de rotation en profondeur. Toutefois, le désavantage amené par la stéréoscopie dans la deuxième session de l'expérience 1 peut difficilement être expliqué par ces dernières études. Cependant, ce renversement s'accorde avec les données de Pasqualotto et Hayward (2009) qui ont montré que la stéréoscopie pouvait augmenter les coûts de rotation lorsque les stimuli étaient familiers. En effet, il semble la familiarité puisse entraîner une extra-spécificité du traitement des éléments picturaux, ce qui pourrait avoir un effet délétère sur les performances en présence d'information stéréoscopique (Pasqualotto & Hayward). Tel que discuté plus haut, les participants à l'Exp. 1 semblent justement avoir développé, à la séance 1, une expertise des stimuli les amenant à se focaliser sur certains éléments picturaux (i.e. 2D) particuliers. Selon l'hypothèse de Pasqualotto et Hayward, Cette expertise particulière serait précisément la cause de l'effet négatif qu'a eu la stéréoscopie normale sur les performances à la séance 2.

Il est aussi intéressant de noter que la stéréoscopie inversée a amené un avantage par rapport à la 2D. Les processus impliqués dans la représentation des formes pourraient donc l'utiliser et en bénéficier. La disparité inversée offre la même information que la disparité normale, mais son signe est inversé par rapport aux informations monoculaires. Suite à l'adaptation à ce renversement, la stéréoscopie inversée semble en mesure de contribuer à l'invariance à l'orientation. Cela est congruent avec les données de Matthews, Hill et Palmisano (2011) qui ont montré que pour l'évaluation de la profondeur de visage, les participants semblaient tenir compte de l'ampleur de la disparité, mais pas de son signe. Une autre possibilité serait que les participants aient donné moins de poids ou aient ignoré l'information monoculaire dans les expériences 2 et 3. D'ailleurs, avec ces stimuli si l'on ignore l'information monoculaire, les formes présentées en stéréo inversée correspondent aux réflexions miroir des formes présentées en stéréo avec des rotations dans le sens inverses.

Implications théoriques générales

Le rôle de l'information binoculaire. La contribution de la stéréoscopie dans la reconnaissance et la constance de forme rapportée dans les articles 1 et 3 est en accord avec les études de Bennett et Vuong (2006), Burke (2005), et Burke et al. (2007) qui avaient montré que l'information binoculaire contribue à la constance de forme. De plus, Julesz (1961, 1970) et d'autres (De Vries, Kappers & Koenderink, 1993; Uttal, Davis & Welke, 1994; Uttal, Davis, Welke & Kakarala, 1988; Vreven, 2006) avaient déjà montré, auparavant, qu'en l'absence de toute autre information de profondeur, la stéréoscopie était suffisante pour reconnaître des formes 3D. Dans les présentes études, l'information de profondeur monoculaire y était plus riche. En effet, dans les expériences des articles 1 et 3, les indices d’ombrage et d’occlusion étaient également disponibles. Les présents résultats indiquent donc que la stéréoscopie contribue à la représentation des formes et ce même avec la disponibilité de riches indices de profondeur monoculaires.

Il est intéressant de noter que d’après la théorie de reconnaissance par composantes (Biederman, 1987; Biederman & Gerhardstein, 1993) et celle de Pizlo (2008), lorsque les formes respectent certaines contraintes (i.e. les conditions d’invariance à l’orientation de Biederman et Gerharstein et les contraintes de simplicité des contours de Pizlo) seuls les contours sont nécessaires à la représentation et à la constance de forme. Les stimuli utilisés dans l’article 3 étaient des trombones tordus et ne répondaient pas à ces différentes conditions. Dans ce contexte, la théorie de la reconnaissance par composantes peut s’accommoder d’une certaine contribution de la stéréoscopie. Par contre, d’après la théorie de Pizlo, l’information binoculaire n’est pas utile à la représentation des formes. Celle-ci est donc contredite par les résultats obtenus à l’article 3 malgré le fait qu’on y ait fait usage de stimuli représentant des trombones tordus. De leur côté, les stimuli utilisés à l’article 1 répondaient aux critères posés par les théories de Biederman et de Pizlo pour que seuls les contours des objets sous-tendent la représentation de leur forme, excluant d’office l’information stéréoscopique. La possible contribution de la stéréoscopie trouvée dans cette expérience falsifie donc ces théories. Il importe néanmoins de souligner que si nous n’avions pas utilisé de condition de stéréoscopie inversée, les conclusions que nous aurions pu tirer auraient été toutes autres. En effet, l’absence de différence entre les performances aux conditions de stéréoscopie normale et 2D aurait été interprétée comme une absence de contribution de l’information binoculaire dans la reconnaissance de ce type de forme. La stéréoscopie inversée est donc une condition utile qui permet de mieux évaluer le rôle de la stéréoscopie.

Finalement, les résultats rapportés l'article 3 montrent que l'impact de la stéréoscopie peut varier en fonction de la familiarité avec les stimuli et en fonction de l'exposition à différents modes de présentation. Le fait que la stéréoscopie ait amené un désavantage dans la deuxième séance de l'expérience 1 est congruent avec les résultats de Pasqualotto et Hayward (2009), qui ont montré que la stéréo peut entraîner une augmentation des coûts de rotation dans la discrimination d'objets familiers.

La disparité inversée. Il est important de souligner que les différentes expériences ont montré différents effets possibles de la stéréo inversée. En effet, elle a eu des effets délétères sur la performance dans l'article 1, mais a été ignorée, a entraîné un certain avantage par rapport à la 2D et a même amené le même avantage que la stéréo normale dans l'article 3. Le fait que la stéréo inversée nuise aux performances n'est pas surprenant puisqu'elle est inconsistante avec l'information monoculaire disponible. Son absence d'impact sur les performances par rapport à la condition de 2D suggère qu'elle a été ignorée et cela appuie les résultats de Norman et Todd, (1995), Saunders et Backus (2006) et and Stevens et al (1991) qui ont montrés que lorsque différentes informations de profondeur sont inconsistantes, la stéréo est ignorée et les autres informations sont utilisées et de Ichikawa et Egusa (1993) qui ont montrés qu'à long terme, moins de poids était donné à la disparité inversée et plus aux informations monoculaires. Finalement, l'avantage qu'elle a amené dans les deux dernières expériences de l'article 3 supporte l'idée avancée par Matthews et Palmisano (2011) selon laquelle on peut utiliser l'amplitude de la disparité indépendamment de son signe. Les processus impliqués dans la reconnaissance de formes pourraient donc tirer avantage de la stéréoscopie inversée, malgré son incongruence avec les informations monoculaires. Cette dernière proposition n'est toutefois pas la seule qui permette d'expliquer l'avantage amené par la stéréo inversée. En effet avec les stimuli utilisés dans l'article 3, si l'on exclu l'information monoculaire,, les formes présentées en 3D inversée correspondaient aux réflexions miroir des formes présentées en 3D, avec des rotations dans des directions opposées. Il est possible que moins de poids ait été donné aux informations monoculaires, ou même qu'elles aient été ignorées, dans les expériences 2 et 3 de l'article 3.

La disparité binoculaire inversée a eu trois types d'effet dans les différentes expériences présentées dans cette thèse. Au meilleur de nos connaissances, aucune étude n'a étudié l'impact de la disparité binoculaire inversée dans la perception des formes, et il est donc difficile d'expliquer pourquoi les effets de cette condition ont été différents entre les différentes expériences présentées ici. Un point intéressant à noter est que les types des stimuli étaient différents entre les différentes expériences, et on peut penser que cela a entrainé des différences dans les effets de la stéréo inversée. Les stimuli de l'expérience 1 sont composés de formes volumétriques. Avec la disparité inversée, les stimuli au complet, ou à tout le moins certaines de leurs composantes paraissent concaves, alors qu'en stéréo normale les composantes ne paraissaient pas concaves. On ne retrouve pas cela avec les stimuli de l'expérience 3. Il est possible que le fait que des parties des formes de l'article 1 soient devenues concaves dans la condition de stéréo inversée ait entrainé un désavantage par rapport à la 2D ou à la stéréo normale. Cela ne pouvait pas se produire avec les stimuli de l'article 3 étant donné leur structure différente et cela pourrait expliquer la différence entre les effets de la stéréo inversée dans les différents articles. Refaire ces expériences avec différents types de stimuli permettrait de tester ces hypothèses.

Finalement, il faut garder en tête que puisque peu ou pas d'études ont évalué l'impact de la stéréo inversée dans la perception et la reconnaissance de forme. Toutes les explications amenées dans les paragraphes précédents sont spéculatives et davantage de recherches devront être conduites pour bien comprendre le rôle de la disparité inversée dans la perception des formes.

Intégration des informations de profondeur binoculaire et monoculaire. Bien qu'il ait souvent été montré que lorsque les indices de profondeur sont incongruents, les plus fiables ont plus de poids et peuvent même outrepasser totalement les autres (Braunstein, Andersen, Rouse & Tittle, 1986; Bülthoff & Mallot, 1988; 1990; Norman & Todd; 1995; Saunders & Backus, 2006; Stevens & Brookes, 1988; Steven, Lee & Brookes, 1991), l'article 2 a montré que pour le jugement du relief, il n'y a pas d'interaction entre l'ombrage et la stéréo, même lorsqu'ils sont incongruents. Cela indique que ces deux types d'information sont intégrés linéairement par les processus impliqués dans la perception du relief. Cela concorde avec le modèle de Landy et al. (1995) qui suppose que les différents indices de profondeur sont intégrés linéairement, ainsi qu'avec les données de Dosher et al. (1986) qui ont montré que la stéréo et la covariance de luminance de proximité étaient aussi intégrées linéairement. L'absence d'interaction entre la stéréo et l'ombrage est aussi en accord avec les résultats de Johnston, Cumming et Parker (1993), qui ont montré que la texture et la stéréoscopie étaient intégrées indépendamment, et de Doorschot, Kappers et Koenderink (2001) qui ont montré qu'il en était de même pour l'ombrage et la stéréoscopie dans l'évaluation locale de surface de formes.

Limites

La manipulation de l'effet du mode de présentation dans l'article 1 était inter-sujets. Il est possible que ceci ait atténué la force des effets observés relativement à une manipulation qui aurait été intra-sujets. Toutefois, il faut souligner que la familiarité avec les stimuli et l’exposition à différents modes de présentation auraient pu altérer les stratégies des participants dans le contexte d’un protocole intra-sujets, tel que démontré dans l’article 3. Une manipulation intra-sujets des conditions d’information de profondeur aurait donc pu empêcher l’observation des différents effets retrouvés dans cette expérience. Il faut aussi noter que la manipulation du mode de présentation était intra-sujet dans l’expérience de l’article 2. Les résultats auraient peut-être été différents si cette manipulation avait été inter-sujet. En effet, puisque dans l’article 3, l’exposition à différent mode de présentation faisait en sorte que la stéréo était consistante avec les informations de profondeur pour le tiers des essais seulement, les participants passaient d’une stratégie consistant à utiliser l’information stéréoscopique à une stratégie consistant à utiliser des détails picturaux invariants à l’orientation. Il est possible que les participants de l’expérience de l’article 2 aient accordé moins de poids à l’information stéréoscopique que si elle avait varié de manière inter-sujet; on pourrait même penser que l’intégration des informations binoculaire et monoculaires n’aurait pas été linéaire dans ce cas et que l’information stéréoscopique aurait eu un véto sur l’ombrage dans le jugement de relief. Il faudrait tester cette possibilité dans une étude future.

Finalement, bien que les différences dans les résultats obtenus dans les expériences présentées ici pourraient être expliquées par des différences entre les types de stimuli utilisés dans chacune, il n’y a pas eu d’analyse de l’information présente avec observateur idéal. Sans ces analyses on ne peut pas conclure que les différences entre les résultats obtenus dans les différents articles sont réellement liées aux différents types de stimuli, ou si elles sont seulement liées à des différences dans l’information présente dans les stimuli. Il est donc difficile de comparer les résultats des trois articles entre eux, mais ce serait intéressant d’utiliser un observateur idéal avec les différents stimuli utilisés ici dans une future étude.

Perspectives futures

Dans des études futures, il serait intéressant d'étudier l'impact des différentes informations de profondeur sur la perception de différents types de formes. Tjan et Legge (1998) ont déjà montré que l'invariance à l'orientation dépendait du type de forme utilisée. Il est fort possible que les effets de la stéréoscopie dans la reconnaissance et la constance de forme soient modulés par le type de forme utilisé. Par exemple, bien que la stéréoscopie normale permette de réduire les coûts de rotation dans l'appariement de trombones tordus par rapport à l'absence de disparité et la stéréoscopie inversée, il est possible, si l'on se fie aux résultats de l'article 1, qu'avec une tâche d'appariement de géons, on ne trouve pas de différence dans les coûts de rotation entre stéréo normale et absence de disparité, mais que la stéréoscopie inversée les augmente. Il serait donc intéressant d'explorer d'avantage les impacts de l'information binoculaire en fonction du type de stimuli utilisé.

Aussi, l'article 2 a montré que l'ombrage et la stéréo contribuent indépendamment à la perception du relief. Il serait intéressant d'évaluer ce qu'il en est pour la constance de forme, et de mesurer comment les informations binoculaire et monoculaires sont intégrées par les différents processus qui sont en jeu. D'ailleurs, l'article présenté dans l'annexe 1 rapporte quelques données amassées sur le sujet au cours des collectes de données pour cette thèse.

Finalement, on sait que dans le système visuel, il y a une dissociation des voies ventrale et dorsale. La voie ventrale est impliquée dans la reconnaissance d'objets et dans le traitement de la texture et des couleurs, alors que la voie dorsale est impliquée dans la localisation spatiale et dans le guidage visuel de l'action (Goodale & Milner, 1992; Goodale et Westwood 2004; Ungerleider & Mishkin, 1982). Malgré cette dissociation, on a trouvé des neurones de la voie ventrale qui étaient sensibles à la disparité binoculaire (Hinkle & Connor, 2001; Uka, Tanaka, Yoshiyama, Kato & Fujita, 2000), des régions de la voie dorsale qui étaient sensibles à la forme 3D créée par la combinaison de différentes informations de profondeur (Welchman, Deubelius, Conrad, Bülthoff and Kourtzi, 2005), et on a trouvé que certaines régions de la voie dorsale étaient sensibles à des formes définies par de la texture (Georgieva, Todd, Peeters & Orban, 2008). Il semble donc que les voies ventrale et dorsale collaborent pour intégrer les différentes informations de profondeur et mener à des représentations des formes efficaces. D'ailleurs, le fait que la stéréo contribue à la perception, à la reconnaissance et à la constance de forme, tel que démontré dans la présente thèse, appuie cette idée. Cependant, il reste à étudier comment cette collaboration s'effectue entre les deux voies.

Conclusion

En bref, les trois articles présentés ici ont permis de montrer que la stéréoscopie contribuait à la reconnaissance et à la constance de forme, mais que cela est fonction du mode de présentation et de la familiarité avec les stimuli. Ils ont aussi montré que la stéréoscopie inversée était une condition utile pour explorer le rôle de l'information binoculaire dans la perception des formes, et que l'information binoculaire et l'ombrage sont intégrés linéairement par les processus sous-tendant la perception du relief.

Bibliographie

Adams, W.J. (2007) A common light-prior for visual search, shape, and reflectance judgements, *Journal of Vision, 7,* 1-7.

Adams, W.J. et Mamassian, P. (2004) Bayesian combination of ambiguous shape cues, *Journal of Vision, 4*, 921-929.

Arguin, M. et Saumier, D. (2004) Independent processing of parts and of their spatial organization in complex visual objects, *Psychological Science, 15*, 629-633.

Bennett, D.J. et Vuong, Q.C. (2006) A stereo advantage in generalizing over changes in viewpoint on object recognition tasks, *Perception and Psychophysics, 68*, 1082-1093.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding, *Psychological Review, 94,* 115-147.

Biederman, I. et Bar, M. (1999) One-shot viewpoint invariance in matching novel objects, *Vision Research, 39,* 2885-2899.

Biederman I, et Gerharstein, P.C. (1993) Recognizing depth-rotated objects: Evidence and conditions for three-dimensional view-point invariance, *Journal of experimental Psychology: Human perception and performance, 19*, 1162-1182.

Biederman, I. et Ju, G. (1988) Surface versus edge-base determinants of visual recognition, *Cognitive Psychology, 20,* 38-64.

Blais, C., Arguin, M., et Marleau, I (2009) Orientation invariance in visual shape perception, *Journal of Vision, 9*, 1-23.

Booth, M.C.A. et Rolls, E.T. (1998) View-invariant representation of familiar objects by neurons in the inferior temporal visual cortex, *Cerebral Cortex, 8*, 510-523.

Braunstein, M.L., Andersen, G.J., Rouse, M.W., et Tittle, J.S. (1986) Recovering viewer-centered depth from disparity, occlusion and velocity gradients, *Perception & Psychophysics, 40*, 216-224.

Bülthoff, H.H. et Mallot, H.A. (1988) Integration of depth modules: stereo and shading, *Journal of Optical Society of America, 5,* 1749-1758.

Burgund, E.D. et Marsolek, C.J. (2000) Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems, *Psychonomic Bulletin & Review, 7,* 480-489.

Burke, D. (2005) Combining disparate views of objects: Viewpoint costs are reduced by stereopsis, *Visual Cognition, 12,* 201-213.

Burke, D., Taubert, J. et Higman, T. (2007) Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces, *Vision Research, 47*, 2164-2169.

Cavina-Pratesi, C., Kentridge, R.W., Heywood C.A. et Milner, A.D. (2010) Separate Processing of texture and form in the ventral stream: Evidence from fMRI and visual agnosia, *Nature Neuroscience, 20*, 433-446.

Chan, M.W., Stevenson, A.K., Li, Y., et Pizlo, Z. (2006) Binocular shape constancy from novel viewa: The role of a priori constraints, *Perception & Psychophysics, 68,* 1124-1139.

Chandrasekaran, C., Canon, V., Dahmen, J.C., Kourtzi, Z., Welchman, A.E. (2007) Neural correlates of disparity-defined shape discrimination in the human brain, *Journal of Neurophysiology, 97*, 1553-1565.

Connor, C.E. (2001) Visual Perception: Sunny side up, Current Biology, 11, R776-R778.

Curran, W., et Johnston, A. (1996) The effect of illuminant position on perceived curvature, *Vision Research, 36(*, 1399-1410.

De Vries, S.C., Kappers, A.M.L., et Koenderink, J.J. (1993) Shape from stereo: A systematic approach using quadratic surfaces, Perception & Psychophysics, 53(, 71-80.

Doorschot, P.C.A., Kappers, A.M.L. et Koenderink, J.J. (2001) The combined influence of binocular disparity and shading on pictorial shape, *Perception and Psychophysics, 63*, 1038-1047.

Dosher, B.A., Sperling, G., Wurst, S. A., (1986) Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure, *Vision research, 26,* 973-990

Edelman S. et Bülthoff H.H. (1992) Orientation dependance in the recognition of familiar and novel views of three-dimensional objects, *Vision Research, 32,* 2385-2400.

Erens, R.G.F., Kappers, A.M.L. et Koenderink, J.J. (1993) Perception of local shape from shading, *Perception and psychophysics, 54*, 145-156.

Fleming, E.W., Torralba, A., et Adelson, E.H. (2004) Specular reflections and the perception of shape, *Journal of Vision, 4*, 798-820.

Gauthier , I., Hayward, W.J., Tarr, M.J., Anderson, A.W., Skudlarski, P. et Gore, J.C. (2002) BOLD activity during mental rotation and viewpoint-dependant object recognition, *Neuron, 34*, 161-171.

Georgieva, S. S., Todd, J.T., Peeters, R. et Orban, G.A. (2008) The extraction of 3D shape from texture and shading in the human brain, *Cerebral Cortex, 18*, 2416- 2438.

Gerardin, P., de Montalembert, M. et Mamassian, P. (2007) Shape from shading: New perspectives from the Polo Mint stimulus, *Journal of Vision, 7*, 1-11.

Gibson, B.M., Lazareva, O.F., Gosselin, F., Shyns, P.G., et Wasserman, E.A. (2007) Nonaccidental properties underlie shape recognition in mammalian and nonmammalian vision, *Current Biology, 14,* 336-340.

Goodale, M.A. et Milner, A.D. (1992) Separate visual pathways for perception and action, *Trends in Neuroscience*, *15*, 20-25.

Goodale, M.A. et Westwood, D.A.(2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action, Current Opinion in Neurobiology, 14, 203-211.

Hayward, W.G. et Tarr, M.J. (1997) Testing Conditions for Viewpoint Invariance in Object Recognition, *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1511-1521.

Hayward, W.G., Zhou, G., Gauthier, I. et Harris, I.M. (2006) Dissociating viewpoint costs in mental rotation and object recognition, *Psychonomic Bulletin & Review, 13*, 820-825.

Hinkle, D.A. et Connor, C.E. (2001) Disparity tuning in macaque area V4, *Neuroreport, 12*, 365-369.

Hummel, J.E.(2001) Complementary solutions to the binding problem in visions: Implication for shape perception and object recognition, *Visual Cognition, 8*, 489-517.

Hummel, J.E. et Biederman, I. (1992) Dynamic Binding in a Neural Network for Shape Recognition, *Psychological Review, 99,* 480-517.

Ichikawa, M., et Egusa, H. (1993) How is depth perception affected by long-term wearing of left-right reversing spactacles?, *Perception, 22*, 971-984.

Ichikawa, M., Egusa, H., Nakatsuka, M., Amano, J., Ueda, T., et Tashiro, T. (2003), Modification of depth and distance perception caused by long-term wearing of left-right reversing spectacles, *Perception, 32,* 131-153 .

Johnston, E.B., Cumming, B.G. et Parker, A.J. (1993) Integration of depth modules: Stereopsis and texture, *Vision Research, 33*, 813-826.

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

Julesz, B. (1960) Binocular depth perception of computer-generated patterns, The Bell System Technical Journal, *39*, 1125-1162.

Julesz, B. (1971) Foundations of cyclopean perception, Chicago: University of Chicago Press.

Khang, B., Koenderinck, J.J., et Kappers, A.M.L. (2007) Shape from shading from images rendered with various surface types and light fields, *Perception, 36*, 1191-1213.

Kleffner, D.A., et Ramachandran, V.S. (1992) On the perception of shape from shading, *Perception and Psychophysics, 52,* 18-36.

Kourtzi, Z. et Kanwisher, N. (2001) Representation of perceived object shape by the human lateral occipital complex, *Science, 293*, 1506-1509.

Landy, M.S., Maloney, L.T., Johnston, E.B., et Young, M. (1995) Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research, 35,* 389 - 412

Leek, E.C. (1998) Effects of stimulus orientation on the identification of common polyoriented objects, *Psychonomic Bulletin & Review, 5(4),* 650-658.

Li, Y., et Pizlo, Z. (2011) Depth cues versus the simplicity principle in 3D shape perception, *Topics in Cognitive Science, 3,* 667-685.

Liu, B. et Todd, J.T. (2004) Perceptual biases in the interpretation of 3D shape from shading, *Vision Research, 44*, 2135-2145.

Liu, Y., Vogels, R. et Orban, G.A. (2004) Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex, *The Journal of Neuroscience, 24,* 3795-3800.

Mamassian, P., et Goutcher, R. (2001) Prior Knowledge on the illumination position, *Cognition, 81*, B1-B9.

Mamassian, P., et Landy, M. S. (2001) Interaction of visual prior constraints, *Vision Research*, *41*, 2653-2668.

Marr, D et Nishihara, H.K. (1978) Representation and recognition: organization of three dimensional shapes, *Proceeding of the Royal Society of London, B, 200*, 269-291.

Matthews, H., Hill, H., et Palmisano, S. (2011) Binocular disparity magnitude affects perceived depth magnitude despite inversion of depth order, *Perception, 40*, 975-988.

McManus, I.C., Buckman, J. et Woolley, E (2004) Is light in pictures presumed to come from the left side?, *Perception, 33*, 1421-1436.

Mingolla, E. et Todd, J.T. (1986) Perception of shape from shading, *Biological Cybernetics, 53,* 137-151.

Norman, J.F. et Todd, T.T. (1995) The perception of 3-D structure from contradictory optical patterns, *Perception & Psychophysics, 57,* 826-834.

Norman, J.F., Todd, J.T. et Orban, G.A. (2004) Perception of three-dimensional shape from specular highlights, deformations of shading and other types of visual information, *Psychological Science, 15*, 565-570.

Palmeri, T. J., Wong, A. C.-N., et Gauthier, I. (2004) Computational approaches to the development of perceptual expertise, *Trends in Cognitive Neuroscience, 8*, 378-386.

Pasqualotto, A. et Hayward, W.G. (2009) A stereo disadvantage for recognizing rotated familiar object, *Psychonomic Bulletin & Review, 16,* 832-838

Pentland A. (1989) Shape Information from Shading: A theory About Human Perception, *Spatial Vision, 4*, 165-182.

Perret,D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietman, J.K., Benson, P.J. et Thomas, S. (1991) Viewer –centered and object-centered coding of heads in the macaque temporal cortex, *Experimental Brain Research, 86*, 159-173.

Pizlo, Z. (2008) 3D Shape: Its Unique Place in Visual Perception, MIT Press, Cambridge.

Pizlo, Z., Li, Y. et Francis, G. (2005) A new look at binoculair stereopsis, *Vision Research 45*, 2244-2255.

Pizlo, Z., Li, Y. et Steinman, R.M. (2008) Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose, *Spatial Vision, 21*, 495-508.

Pizlo, Z. et Stevenson, A. (1999) Shape constancy from novel views, *Perception & Psychophysics, 61,* 1299-1307.

Poggio, T. et Edelman, S. (1990) A network that learns to recognize 3D objects, *Nature, 343*, 263-266.

Ramachandran, V.S., (1988) Perception of shape from shading, *Nature, 331*, 163-166.

Saunders, J.A. et Backus, B.T. (2006) Perception of surface slant from oriented textures, *Journal of Vision, 6*, 882-897.

Shepard, R.N. et Metzler, J. (1971) Mental Rotation of Three-Dimensional Objects, *Science, 171,* 701-703.

Shimojo, S. et Nakajma, Y. (1981) Adaptation to the reversal of binocular depth cues: Effects of wearing left-right reversing spectacles on stereoscopic depth perception, *Perception, 10*, 391-402.

Stankiewicz, B. (2002) Empirical Evidence for independent dimensions in the visual representation of three-dimensional shape, *Journal of experimental psychology: Human perception and performance, 28*, 913-932.

Stevens, K.A., et Brookes, A. (1988) Integrating stereopsis with monocular interpretations of planar surfaces, *Vision Research, 28*, 371-386.

Stevens, K.A., Lees, M., et Brookes, A. (1991) Combining binocular and monocular features, *Perception, 20*, 425-440.

Stone, J.V., Buckley, D. et Moger, F.A. (2000) Determinants of object recognition, Vision Research, 40, 2723-2736.

Sun, J., et Perona, P. (1998) Where is the sun?, *Nature Neuroscience, 1,* 183-184.

Tarr, M.J. (1995) Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects, *Psychonomic Bulletin & Review, 2,* 55-82.

Tarr, M.J. et Pinker, S. (1989) Mental Rotation and Orientation-Dependence in Shape Recognition, *Cognitive Psychology, 21,* 233-282.

Tarr, M.J., Williams, P., Hayward, W.G. et Gauthier, I. (1998.) Three-dimensional object recognition is viewpoint dependent, *Nature Neuroscience, 1,* 275-277.

Tjan, B.S. et Legge, G.E. (1998) The viewpoint complexity of an object-recognition task, *Vision Research, 38,* 2335-2350.

Todd, J.T., et Norman, J.F. (2003) The visual perception of 3-D shape from multiple cues: are observers capable of perceiving metric structure? *Perception & Psychophysics, 65,* 31-47

Todd, T.T., Norman, J.F., Koenderink, J.J., et Kappers, A.M.L. (1997) Effects of texture, illumination and surface reflectance on stereoscopic shape perception, *Perception, 26*, 806-822.

Uka, T., Tanaka, H., Yoshiyama, K., Kato, M.et Fujita, I. (2000) Disparity selectivity of neurons in monkey inferior temporal cortex, *Journal of Neurophysiology, 84*, 120-132.

Ullman, S. (1989) Aligning pictorial descriptions: An approach to object recognition, *Cognition, 32*, 193-254.

Ullman, S. et Basri, R. (1991) Recognition by linear combinations of models, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,* 992-1006.

Ungerleider, L.G., et Mishkin, M. (1982) Two cortical systems, *Analysis of Visual Behavior,* Cambridge, MA, MIT Press, 549-586.

Uttal, W.R., Davis, N.D., et Wlke, C. (1994) Stereoscopic perception with brief exposures, Perception & Psychophysic, *56,* 599-604.

Uttal, W.R., Davis, N.S., Wleke, C., et Kakarala, R. (1988) The reconstruction of static visual forms from sparse dotted samples, Perception & Psychophysics, *43*, 223- 240.

Vreven, D. (2006) 3D shape discrimination using relative disparity derivatives, *Vision Research, 46*, 4181-4192.

Vogels, R., Biederman, I., Bar, M. et Lorincz, A. (2001) Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences, *Journal of Cognitive Neuroscience, 13,* 444-453.

Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H. et Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortx, *Nature Neuroscience, 8,* 820-827.

Wheatsone, C. (1852), Contributions to the physiology of vision part II. On some remarkable, and hitherto unobserved, phaenomena of binocular vision, *Abstract of the papers communicated to the Royal Society of London, 6*.

Wright, M. et Ledgeway, T. (2004) Interaction between luminance gratings and disparity gratings, *Spatial Vision, 17*, 51-74.

Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z. et Connor, C.E. (2008) A neural code for three-dimensional object shape in macaque infero temporal cortex, Nature Neuroscience, 11, 1352-1360.

Yellott, J.I. et Kaiwi, J.L. (1979) Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective, *Perception, 8*, 135-142.

ANNEXE I

Voici un article qui comporte d’autres données qui ont été amassées au long de mon doctorat. Cette étude évalue l’impact, l’intégration et l’interaction des informations de profondeur binoculaire et monoculaires dans la constance de forme.

THE CONTRIBUTION OF BINOCULAR AND MONOCULAR DEPTH CUES TO SHAPE CONSTANCY

Mercédès Aubin & Martin Arguin

Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de psychologie, Université de Montréal

Contribution des coauteurs :

M.Aubin., & M. Arguin. : Conception du projet

M. Aubin. : Recrutement des participants

M. Aubin. : Collecte des données

M. Aubin. : Analyse des résultats

M. Aubin. : Rédaction de l’article

M. Aubin. : Interprétation des résultats

M. Aubin, & M. Arguin : Révision du manuscrit

M. Arguin : Financement

Abstract

The contributions of stereoscopic and monocular depth information in supporting shape constancy were investigated using a task of sequential matching task with depth rotated stimuli shaped like bent paperclips. The stimuli were rotated between 0 and 80 degrees around the vertical axis, and they could be presented without stereopsis, with normal stereopsis or with reversed stereopsis. The monocular information defining the stimuli ranged from silhouettes only to the combination of shading, texture and specular reflections. The factors through which the depth cues available were manipulated were between-subjects. The rotation costs on performance were weaker with the stereoscopic presentation than with either the 2D or reversed stereoscopic displays. Also the rotation costs were weaker when stimuli were defined by shading, than by only silhouettes, texture or specular reflections. Adding more monocular cues to the shaded stimuli did not further decrease the rotation effect. The effects of display mode and monocular information did not interact. The results indicate that stereopsis and shading seem to contribute independently to shape perception, and that shading is more important than texture and specular reflections to achieve shape constancy with the class of stimuli used in the present study.

Introduction

Understanding the representation of three-dimensional (3-D) objects is fundamental to theories of vision (e.g., Arguin & Saumier, 2005; Biederman, 1987; Edelman, 1999; Hummel, 2001; Hummel & Biederman, 1992; Leek, Reppa & Arguin, 2005; Marr & Nishihara, 1978). One unresolved issue concerns the contribution of the binocular and monocular depth cues for recognition and shape constancy.

According to current structural theories (Biederman, 1987; Biederman & Bar, 1999; Biederman & Gerhardstein, 1993; Hummel & Biederman, 1992; Marr, 1982; Marr & Nishihara, 1978; Peissig, Wasserman, Young, & Biederman, 2002; Pentland, 1985), internal shape representations encode the 3D structure of objects in terms of a collection of volumetric primitives and their connectedness. Biederman proposed that the non-accidental properties, which are 3D shape properties that are shared across different viewpoints of an object, would be the basis for its recognition. Thus, in accordance to this view, as long as the non-accidental properties remain visible, shape representations should be completely orientation invariant. This theory suggests that the implicit 3D cues provided by the processing of edges are sufficient to create a 3D shape representation allowing shape constancy. Other kinds of depth cues such as texture, shading or stereopsis are assumed to not contribute. Congruently, Biederman & Ju (1988) showed that object denomination was as fast for line drawn objects as for photographs.

View-based theories, which constitute another major class of theories, suggest that shape representations exclusively code information about 2D views. This necessarily implies that the recognition of an object in a novel orientation requires a special process of alignment, normalisation and/or interpolation to match it to its stored 2D representation (Hayward & Tarr, 1997; Tarr & Bülthoff, 1995; Tarr, Bülthoff, Zabinski, & Blanz, 1997; Tarr & Pinker, 1989, 1990, 1991; Tarr, Williams, Hayward, & Gauthier, 1998). Proponents of view-based theories have remained ambiguous about the role of the different depth cues, including stereopsis, in shape representation.

Proposing a different theory, Pizlo and colleagues (Chan, Stevenson, Li & Pizlo, 2006; Pizlo, 2008; Pizlo, Li & Francis, 2005; Pizlo, Li, Steinman, 2008; Pizlo & Stevenson, 1999) argue that shape representations and shape constancy are based on a priori monocular simplicity constraints. Consequently, binocular depth cues are neither necessary nor sufficient for shape constancy. In support of this view, Pizlo et al. reported that when monocular simplicity constraints and stereoscopic depth cues are inconsistent, binocular cues are ignored. Furthermore, the performances of models of monocular shape representation in a sequential matching task correlate with human monocular and binocular performances while the binocular models performances do not (Chan et al., 2006). However, a more recent study of Li and Pizlo (2011) supports the view that edges and stereopsis are useful to achieve shape constancy. This is consistent with reports that indicate that stereopsis reduces rotation costs in different matching and recognition tasks with stimulus classes for which human beings fail to show complete orientation invariance, i.e. faces, bent paper clip, tube like shapes (Bennett & Vuong, 2006; Burke, 2005; Burke, Taubert, & Higman, 2007).

To further explore the contribution of stereo to shape perception, reversed stereo appears as a relevant exposure condition that has been poorly explored so far. Thus, the studies that have examined the effect of reversed stereopsis on visual perception (Ichikawa & Egusa, 1993; Ichikawa, Egusa, Nakatsuka, Amano, Ueoa & Tashiro, 2003;

Shimojo & Nakajame, 1981; Wheatsone, 1852; Yellot & Kaiwi, 1979) have focused on the adaptation effect resulting from the long term wearing of right-left reversing spectacles which reversed binocular disparities. This adaptation effect led to depth inversion once the spectacles were removed and in an alteration of the weight of the different depth cues (Ichikawa & Egusa, 1993). Indeed, participants ended up ignoring binocular disparity altogether and using occlusion and linear perspective to a greater degree than before adaptation to make depth judgements.

Few studies have been conducted so far to determine the impact of reversed stereopsis on shape constancy and shape recognition. Such a study is relevant however, because if the visual system uses stereopsis to achieve shape constancy, reversed disparity should alter performances compared to normal disparity or the absence of disparity since the information carried by stereopsis would be inconsistent with monocular depth cues such as shading, linear perspective, edges, occlusion, shading, texture, and specular reflections. Aubin and Arguin (submitted) have shown that reversed stereopsis has a deleterious effect on shape recognition performances. Unpublished results by Aubin and Arguin also show that, relative to normal stereo, reversed stereo increases rotation costs in a task of sequential matching of bent paperclips.

Monocular depth cues also have an impact on shape perception. Indeed cues like texture and specular highlights can affect the perception of shapes displayed with stereopsis (Todd, Norman, Koenderinck & Kappers, 1997). These cues allow a more accurate evaluation of the slant of local surfaces than stereo or shading alone (Todd et al., 1997). However, Norman, Todd and Orban (2004) showed that in the discrimination of rock-like blobs, performance was better with the depth cues of specular highlights, shading or stereo than with texture and contour or contour alone. Specular highlights led to the best performance of all. With respect to monocular depth cues, it is relevant to underline that texture and specular highlights are complementary sources of information (Fleming, Torralba & Anderson, 2004). Indeed, the foreshortening of texture is a function of the surface orientation while the foreshortening of specular highlights is a function of the surface curvature.

Not only texture contributes to surface and shape evaluation, it also seems capable of supporting viewpoint invariance at particular intermediate or high-level stages of visual shape processing (Blais, Arguin & Marleau, 2009). In contrast however, Li and Pizlo (2011) suggest that shading and texture are not important to achieve shape constancy, contrarily to edges and binocular disparity. Nonetheless, no study has yet evaluated the role of specular highlights in shape constancy, nor the interaction between binocular and monocular depth cues in shape constancy using a condition of reversed stereopsis.

There is evidence indicating that depth cues are weighted as a function of their reliability and this varies according to context. For instance, texture is more reliable to evaluate large than small slant. Stereopsis is more reliable with small slant but at short viewing distances, and the magnitude of the slant size effect is modulated by viewing distance (Hillis, Watt, Landy & Banks, 2004; Knill & Saunders, 2003). These two cues seem to be optimally integrated and are given weights proportional to their reliability when it comes to slant discrimination and judgements (Hillis et al.; Knill & Saunders). When these cues are in conflict however, each receives different weights depending on which is the most informative for slant evaluation (Saunders & Backus, 2006). Norman and Todd (1995) found that for the perception of surface corrugation in depth, when stereo and motion contradict each other, the modality showing the more effective surface curvature direction (horizontal or vertical) is perceived and the other suppressed. Furthermore, some studies have shown that when stereopsis and monocular cues (occlusion and velocity or texture gradients) are inconsistent, the monocular cues override stereopsis (Braunstein, Andersen, Rouse & Tittle, 1986; Stevens & Brookes, 1988). However, Bütlhoff and Mallot (1988; 1990) report that for local surface evaluation, if depth cues are in conflict, edge-based stereo overrides disparate shading and non-disparate shading. Furthermore, disparate shading inhibits non-disparate shading.

While a number of studies indicate that when different depth cues are available, some may override the others, the more common case is cue integration. Landy and colleagues have proposed the modified weak fusion (MWF) model as a general account of how depth cues are integrated (Landy, Maloney, Johnston, & Young, 1995). According to the MWF model, depth cues are weighted according to their reliability, availability, and consistency, and they are typically integrated linearly. The model can accommodate nonlinearities however, such as one cue vetoing another, in particular cases where cues are inconsistent or unreliable. In the same vein, Aubin and Arguin (2014) showed that shading and stereopsis are integrated linearly in relief perception. Dosher, Sperling and Wurst (1986) demonstrated that in the perception of 3D structure, stereo and proximity luminance covariance (i.e. the increase in edge intensity as a function of the proximity to the observer) are integrated linearly. Oruç, Maloney and Landy (2003) showed that for some participants, linear perspective and texture were combined linearly in slant evaluation. Svarverud, Gilson and Glennerster (2010) demonstrated that texture, stereo and motion parallax are combined linearly in 3D location judgement. Ernst and Banks (2002) and Helbig and Ernst (2007) showed that visual and haptic information are integrated linearly in shape and size evaluation. Finally, Watt, Akeley, Ernst and Banks (2005) demonstrated that focus cues, stereopsis and linear perspective are integrated linearly in slant perception. It is important to underline that no study has yet evaluated how the different depth cues are integrated in shape constancy.

The goal of the present study is to determine if stereopsis and monocular depth cues (texture, shading and specular highlights) contribute to shape constancy for non-familiar shapes for which humans do not show complete viewpoint invariance, and to determine how these cues are integrated. The method for the current experiment is similar to that of Burke (2005), who used a sequential matching task with bent paperclips as stimuli. We replicated the stimuli of Burke, but enlarged them so that we could manipulate their surface properties. In one condition, objects were presented as silhouettes and were thus deprived of monocular depth cues. We also used six other conditions that differed according to the depth cues available: shading only, specular highlights only, texture only, specular highlights plus shading, texture plus shading, and specular highlights plus texture and shading (see Fig. 1). In addition, the stimuli from any of these conditions could be presented with either no disparity, normal binocular disparity or reversed binocular disparity. The variations of binocular and monocular depth cues were between-subject. Based on previous studies conducted in our laboratory (Aubin & Arguin, submitted), it was expected that rotation costs would be lower with stereoscopic displays than with either reversed stereo or null disparity. In addition, an interaction between monocular and binocular depth cues may be expected given that previous studies have shown that when depth cues are inconsistent, the most informative cue overrides the others (Bülthoff & Mallot, 1988; 1990; Norman & Todd; 1995; Saunders & Backus, 2006; Steven et al, 1991), and it is often stereo that ends up being ignored (Braunstein & al, 1986; Stevens & Brookes, 1988). Alternatively, based on the modified weak fusion (MWF) model of depth cue integration (Landy et al., 1995) as well as other studies supporting its predictions (Aubin & Arguin, 2014; Dosher et al., 1986; Ernst & Banks, 2002; Helbig & Ernst, 2007; Oruç et al., 2003; Svarverud et al., 2010; Watt et al., 2010) a linear integration of the different depth cues should be expected.

Insert Fig. 1 near here

Methods

Participants

Twenty-one groups of twelve participants aged between 18 and 39 years old took part in the experiment. All were naive as to the purpose of the experiment, were neurologically intact, and had normal or corrected visual acuity. No particular selection was applied with respect to gender, manual dominance, or level of education.

Material and Stimuli

The stimuli were presented on a 16-inch Compaq monitor with 1024 x 768 pixels resolution. The experiment was run on a Pentium 4 computer, and its progress and registration of the observer’s performance were controlled by the E-Prime software. Participants responded by pressing the buttons of a Serial Response Box by Psychology Software Tools.

The stimuli are shown in Fig. 1. They were created using the 3D Studio MAX® program from Autodesk Media & Entertainment (CA, USA). They are four bent paperclip like shapes with a spatial extent of 28 x 33 degrees of visual angle presented at a viewing distance of 35cm. Each bent paperclip was presented at five possible viewpoints (0, 20, 40, 60 or 80 degrees of rotation around the vertical axis; see Fig. 2.) Seven versions of the stimuli were created by varying the surface information they depicted; the items were either silhouettes, or they comprised monocular depth cues of shading only, specular highlights only, checkered texture only, specular highlights and shading, texture and shading, or a combination of specular highlights, texture and shading (see Fig. 1). Rotations of 10.6 degrees around the vertical axis were applied to the stimuli to create distinct views for the left and right eyes for the stereo and reversed stereo conditions. The images were then fused to produce red/cyan anaglyphs. For the 2D display condition, stimuli with no disparity were fused to make their preparation and viewing conditions (all stimuli were viewed with red/cyan glasses, which participants wore throughout the experiment) identical across conditions.

____

Insert Fig. 2 near here

____

Procedure

Participants performed a sequential matching task with a 2 (response: same or different) x 5 (orientation: 0^o, 20^o, 40^o, 60^o, 80^o) x 3 (display mode: no disparity, disparity, reversed disparity) x 7 (monocular cues: external edges, shading, texture, specular highlights, shading and texture, shading and specular highlights, shading, texture and specular highlights) mixed measures experimental design. The variation of monocular depths cues and display mode were between-subject. The viewing distance of 35 cm was maintained constant using a chin rest. The background of the screen was neutral grey. At the beginning of each trial, a fixation cross was displayed for 500 ms, followed by the first stimulus for 2500 ms, followed by a mask of white noise for 500 ms and then by the second stimulus that lasted until the participant's response, or up to a 5000 ms delay. All inter-stimulus intervals were of 0 ms. Visual feedback ("incorrect" written in the middle of the screen) of 1000 ms was given when the participant made an error. There was a 500 ms delay between each trial. Participants were instructed to respond as fast and accurately as possible. Half of the participants indicated that the stimuli were the same by pressing the right button of the response box with the right index finger and that the stimuli were different using the left button pressed the left index finger. These assignments were reversed for the other half of participants. For each object, there were 25 possible combinations of orientations for "same" pairs (e.g. Stimulus 1 at 0^o and Stimulus 1 at 0^o, Stimulus 1 at 0^o and Stimulus 1 at 20^o, and so on) and each combination was presented twice with a switch in stimulus order across repetitions. This (four objects x 25 combinations of orientations x 2 presentations) gives a total of 200 "same" trials for each participants. The 200 "different" trials had the same structure, but the second stimulus was replaced by a different object which was randomly selected with the constraint that each object and orientation occurred in equal numbers. The complete experiment comprised a total of 400 trials, which were divided in two blocks of 200 trials each. The overall set of trials for each participant was distributed randomly across the two blocks.

Results

Data from 3 subjects (1.2%) were rejected because their response times (RTs) or error rates (ERs) exceeded 2.5 standard deviations from the group means. Response times that were more than 2.5 standard deviations away from a participant's average for a given condition were eliminated, resulting in the exclusion of a total of 2413 data points for the complete experiment (2.39 % of trials). The error trials were also excluded from the RTs analyses (14.0% of trials). The data showed no speed-accuracy trade-off in any group since the correlations between corrects RTs and error rates were positive (r ranging from +.41 to +.89, $p < .08$). Table 1 presents the results of the linear regression analyses of correct RTs as well as of ERs as a function of the degree of rotation for each group.

Insert Table 1 near here

RTs analysis. Data analyses were conducted for "same" trials only since orientation differences between different objects are irrelevant.

A three-way mixed-design ANOVA including the factors of rotation (0, 20, 40, 60 and 80 degrees of rotation), display mode (2D, stereo, reversed stereo) and monocular information (silhouettes, shading, specular highlights, texture, shading and specular highlights, shading and texture and shading, specular highlights and texture) was carried out on correct RTs. A main effect of rotation was obtained (*F* (4, 912) = 402.28, *p* < .001), indicating that RTs increased with increased rotation (see Table 1). The results show no main effect of display mode (*F* < 1) or of monocular information (*F* < 1). A significant interaction of rotation x display mode was obtained (*F* (8, 912) = 2.40, *p* < .05; see Fig. 3). Pairwise contrasts between display conditions showed that the rotation cost was significantly weaker in the 3D than the 2D (*F* (4, 656) = 13.59, *p* < .001) or the reversed 3D conditions (marginal effect, *F* (4, 660) = 2.02, *p* = .07), with no difference between the latter (*F* < 1). Monocular information (Fig. 4) did not interact with any other factor and no other significant effect was found (all *F's* < 1).

Insert Figs. 3 and 4 near here

ERs analysis

A three-way mixed-design ANOVA including the factors of training, rotation, display mode and monocular information was carried out on ERs. Main effects of rotation (*F* (4, 912) = 485.65, *p* < .001), display mode (*F* (2, 228) = 10.23, *p* <. 001) and monocular information (*F* (6, 228) = 4.03, *p* <.005) were obtained. The effect of rotation indicates that ERs increased with increased rotation (see Table 1). They also indicate that ERs varied as a function of display mode and monocular information, for which the relevant analyses are described below.

These main effects were qualified by the two-way interaction of rotation x display mode (*F* (8, 912) = 7.25, *p* < .001; see Fig. 5). Pairwise contrasts between display conditions indicate lower ERs with 3D displays than with 2D (F (4, 656) = 15.77, $p < .001$), and reversed 3D (F (4, 656) = 14.48, $p < .001$) displays. They also show a weaker rotation cost weaker with 3D than 2D (F (4, 656) = 12.37, $p < .001$) or reversed 3D displays (F (4, 656) = 9.75, $p < .001$), but no difference between the latter conditions ($F < 1$).

Insert Fig. 5 near here

The results also indicate a significant interaction of rotation x monocular information (F (24, 912) = 3.60, $p < .001$; Fig. 6). The outcome of the pairwise contrasts between the different monocular cue conditions is fully summarized in Table 2.

Insert Fig. 6 and Table 2 near here

Discussion

Both RTs and ERs, increased with an increase in the degree of rotation between the stimuli to be compared. Display conditions also had an important impact on performances and interacted with rotation. Thus, rotation costs were weaker in the normal stereo than the non disparity display mode (for both RTs and ERs), and the reversed stereo condition (ERs only).

Monocular information also had an impact on the ERs and interacted with rotation. Indeed, ERs and rotation costs on ERs were weaker when the stimuli were defined by shading than when they were defined by silhouettes, texture, and specular highlights, while the results did not differ between the latter conditions. Somewhat surprisingly, the combination of shading with other individual monocular depth cues had no impact relative to these other cues alone, except for two cases: shading combined with texture led to weaker rotation costs than texture alone; shading combined with specular reflections led to greater rotation costs than shading alone. However, the combination of shading with texture and specular reflections did lead to significant benefits relative to silhouettes, specular reflections or texture alone (on ERs and rotation costs on ERs) and relative to the combination of shading with specular highlights (on rotation costs only). Finally, the effects of display mode and monocular information did not interact in any way in the present experiment.

Shape constancy and stereopsis

The rotation costs observed were expected. Indeed, any of the theories described in the introduction would predict these effects since bent paperclips do not meet the rotation invariance conditions of Biederman and Gerharstein (1993) or the simplicity constraints of Pizlo and Stevenson (1999). By their very nature, view-based theories also predicted significant rotation effects.

The reduced rotation cost in the normal stereo condition compared to the 2D condition indicates that stereopsis contributes to shape constancy. This is consistent with the previous results of Bennett and Vuong (2006), Burke (2005), Burke et al. (2007), and

Li and Pizlo (2011). A more novel finding from the present study concerns our assessment of the effect of reversed stereo on performance. Rotation costs did not differ between zero disparity displays and reversed 3D, even while significant benefits of normal stereo were observed relative to these conditions. Thus, even though the general constraints of the task used were appropriate to elicit a favorable use of normal binocular disparity information, reversed binocular disparities failed to affect performance. This supports the idea that when stereopsis is inconsistent with other depth cues or information about the shape, it is ignored. It has already been shown that when depth cues are inconsistent, more weight is given to that which is more informative, and, less weight to the less informative which might even be ignored (Braunstein & al, 1986; Bülthoff & Mallot, 1988; 1990; Norman & Todd; 1995; Saunders & Backus, 2006; Stevens & Brookes, 1988; Steven & al, 1991).

The impact of monocular information in shape constancy

Shading. Our results show that the ease with which shape constancy is achieved is affected by the nature of the monocular depth information available and shading stands out as the most crucial cue. This may appear unsurprising given that it has long been known that shading is important for the interpretation of depth in shape perception (Cavanagh & Leclerc, 1989; Mingolla & Todd, 1986; Ramachandran, 1988). However other studies have cast doubt on the usefulness of shading by showing that on its own, it may be insufficient to evaluate shapes properly (Erens, Kappers & Koenderink, 1993) and even suggesting that it does not contribute to shape constancy (Li & Pizlo, 2011). The present findings clearly stand in contradiction to the latter view by showing that shading leads to a substantial reduction of rotation costs not only relative to silhouettes, but also compared to the relatively potent monocular depth cues of texture and specular reflections.

Texture. In contrast to shading, texture failed to improve performance compared to silhouettes. In fact, the only gain to which texture may possibly have contributed is when it was combined together with shading and specular highlights (see Table 2), but then again, this gain was not significantly greater than that offered by shading alone. These observations are congruent with those of Norman et al. (2007), who reported that texture is less efficient than shading for surface evaluation. They are also consistent with the findings of Li and Pizlo (2011), who showed that texture is not useful to achieve shape constancy. However, the present results stand in apparent conflict with those of Blais et al. (2009), who showed that texture results in the viewpoint invariance of representations at particular intermediate or high levels of shape processing (see Arguin & Saumier, 2000). Our results are also inconsistent with the observation that texture significantly contributes to the evaluation of the depth orientation of object surfaces (Todd & al, 1997; Saunders & Backus, 2006). A relevant distinction in relation to these inconsistencies is the observation that texture is a better cue for the perception of surface orientation and shading a better cue for shape perception (Bülthoff & Mallot, 1990). In addition, it is worth noting that experiments showing benefits from texture used stimuli like blobs, planar surfaces, and deformed ellipsoids. In contrast to these classes of shapes, tubular objects such as used here have no large parts of surfaces where texture gradients can be highly visible and this may have played in its poor contribution to shape constancy here.

Specular highlights. Specular highlights seem to not contribute to shape constancy any more than texture. In fact, we even found that specular highlight plus shading led to slightly greater rotation costs than shading alone. This should not be interpreted in terms of a negative impact of specular highlights however, since when this cue was combined with shading and texture, ERs and rotation costs on ERs were weaker than in several other conditions (Table 2), even though this cue combination brought no significant gain relative to shading alone. Rather, our results suggest that specular highlights do not contribute to the processes leading to the rotation invariance of shapes like bent paperclips.

To the best of our knowledge, the present study is the first to evaluate the impact of specular highlights on shape constancy. Our results remain somewhat surprising nevertheless considering that specular highlights lead to better performance than shading or texture alone in shape discrimination (Norman et al. 2004), that foreshortening of these highlights is a function of the surface curvature (Fleming et al., 2004), and that bent paper clips are especially rich in curves.

The independence of binocular and monocular depth cues in shape representation.

It is interesting to note that we found no interaction between the effect of display mode and monocular cues. This indicates that binocular disparity information had not impact on the effect monocular depth cues, and vice-versa. This observation indicates that stereopsis and monocular cues are integrated linearly, consistently with the MWF model (Landy et al., 1995), which postulates that depth cues are integrated linearly. This observation is also congruent with the data of Aubin and Arguin (2014), Dosher et al.

(1986), Ernst and Banks (2002), Helbig and Ernst (2007) Oruç et al. (2003), Svarverud et al. (2010), and Watt et al. (2010) who showed that depth cues were integrated linearly in relief, 3D structure, slant, shape and size perception. However, they disagree with the results of other previous studies showing that the more reliable cues are given more weight or that they may even override the less informative cues (Bülthoff & Mallot, 1988; 1990; Norman & Todd; 1995; Saunders & Backus, 2006; Steven, Lee & Brookes, 1991) – in such cases it is monocular cues which most often override binocular information (Braunstein, Andersen, Rouse & Tittle, 1986; Stevens & Brookes, 1988). We propose that these different outcomes are a function of differences in the nature of the tasks required of the participants. Whereas the present task required the discrimination of rotated shapes, those used in experiments showing a non-linear integration of depth cues used local surface depth evaluation (Bülthoff & Mallot), slant evaluation (Saunders & Backus) or surface evaluation (Norman & Todd, 1995; Stevens & Brookes, 1988; Steven et al. 1991). The markedly different constraints of the tasks and the nature of the stimulus interpretation required most likely led to the difference in how depth information was used.

Future studies

It is widely recognized that the ventral pathway processes visual objects, forms, texture and color, and that the dorsal pathway is implicated in spatial localisation and in the visual guidance of action (Goodale & Milner, 1992; Goodale et Westwood 2004; Ungerleider & Mishkin, 1982). However it has been found that some neurons in the ventral pathway show some sensitivity to binocular disparity (Hinkle & Connor, 2001;

Uka, Tanaka, Yoshiyama, Kato & Fujita, 2000), that V5, part of the dorsal pathway, is sensitive to 3D shapes created by the combination of different depth cues (Welchman, Deubelius, Conrad, Bülthoff & Kourtzi, 2005), and that the intra-parietal sulcus, also part of the dorsal pathway, is sensitive to shape defined by texture but not by shading (Georgieva, Todd, Peeters & Orban, 2008). These results are consistent with the contribution of stereopsis and monocular depth cues to shape constancy that we measured here since both pathways seem to integrate the depth cues to achieve shape representation. It is already known that the recognition of rotated objects is associated with an increase in the activity of some ventral areas as a function of rotation magnitude (Gauthier et al, 2002), but it would be interesting to evaluate how stereo and monocular depth cues may impact on that activity.

Summary

The aim of the present experiment was to evaluate the impact of binocular and monocular depth cues and their interaction on shape constancy using shapes for which humans do not present complete orientation invariance. The results show that stereopsis and shading contribute to shape constancy, whereas the availability of texture and specular highlight information failed to improve performances. Finally, there is no interaction between the effects of display mode and monocular cues, indicating that these sources of information are integrated linearly.

References

Aubin, M., & Arguin, M. (2014) Stereo and shading contribute independently to shape convexity/concavity discrimination, *Perception*.

Aubin, M. & Arguin, M. (submitted) The conditional contribution of stereopsis to shape constancy, *Quarterly Journal of Experimental Psychology* .

Bennett, D.J., & Vuong, Q.C. (2006) A stereo advantage in generalizing over changes in viewpoint on object recognition tasks, *Perception and Psychophysics, 68*, 1082-1093.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding, *Psychological Review, 94*, 115-147.

Biederman, I., & Ju, G. (1988) Surface versus edge-base determinants of visual recognition, *Cognitive Psychology, 20(1),* 38-64.

Braunstein, M.L., Andersen, G.J., Rouse, M.W., & Tittle, J.S. (1986) Recovering viewer-centered depth from disparity, occlusion and velocity gradients, *Perception & Psychophysics, 40*, 216-224.

Bülthoff, H.H., & Mallot, H.A. (1988) Integration of depth modules: stereo and shading, *Journal of Optical Society of America, 5,* 1749-1758.

Bülthoff, H.H. & Mallot, H.A. (1990) Integration of stereo, shading and texture, AI and the Eye, Jonh Wiley & Sons Ltd,

Burke, D. (2005) Combining disparate views of objects: Viewpoint costs are reduced by stereopsis, *Visual Cognition, 12*, 201-213.

Burke, D., Taubert, J., & Higman, T. (2007) Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces, *Vision Research, 47*, 2164-2169.

Cavanagh, P., & Leclerc, Y.G. (1989) Shape from shadows, *Journal of Experimental Psychology: Human Perception and Performance, 15*, 3-27.

Chan, M.W., Stevenson, A.K., Li, Y., & Pizlo, Z. (2006) Binocular shape constancy from novel views: The role of a priori constraints, *Perception & Psychophysics, 68*, 1124-1139.

Edelman S., & Bülthoff H.H. (1992) Orientation Dependance in the Recognition of Familiar and Novel Views of Three-Dimensional Objects, *Vision Research, 32*, 2385-2400.

Erens, R.G.F., Kappers, A.M.L. & Koenderink, J.J., (1993) Perception of local shape from shading, *Perception and psychophysics, 54*, 145-156.

Ernts, M.O., & Banks, M.S., (2002) Humans integrate visual and haptic information in a statistically optimal fashion, *Nature, 415*, 429-433.

Gauthier , I., Hayward, W.J., Tarr, M.J., Anderson, A.W., Skudlarski, P. & Gore, J.C. (2002) BOLD activity during mental rotation and viewpoint-dependant object recognition, *Neuron, 34*, 161-171.

Georgieva, S. S., Todd, J.T., Peeters, R. & Orban, G.A. (2008) The extraction of 3D shape from texture and shading in the human brain, *Cerebral Cortex, 18*, 2416-2438.

Goodale, M.A. & Milner, A.D. (1992) Separate visual pathways for perception and action, *Trends in Neuroscience*, 15, 20-25.

Goodale, M.A. & Westwood, D.A.(2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action, *Current Opinion in Neurobiology, 14*, 203-211.

Hayward, W.G., & Tarr, M.J. (1997) Testing Conditions for Viewpoint Invariance in Object Recognition, *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511-1521.

Helbig, H.B., Ernst, M.O. (2007) Optimal integration of shape information from vision and touch, *Experimental Brain Research, 179*, 595-606.

Hinkle, D.A. & Connor, C.E. (2001) Disparity tuning in macaque area V4, *Neuroreport, 12*, 365-369.

Ichikawa, M. & Egusa, H. (1993) How is depth perception affected by long-term wearing of left-right reversing spactacles?, Perception, 22, 971-984.

Ichikawa, M., Eguase, H., Nakatsuka, M., Amano, J., Ueda, T., Tashiro, T. (2003), Modification of depth and distance perception caused by long-term wearing of left-right reversing spectacles, *Perception, 32,* 131-153.

Knill, D.C. & Saunders, J.A. (2003) Do human optimally integrate stereo and texture information for judgement of surface slants?, *Vision Research, 43*, 2536-2558.

Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. (1995) Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research, 35*, 389-412.

Li, Y., & Pizlo, Z. (2011) Depth cues versus the simplicity principle in 3D shape perception, *Topics in Cognitive Science, 3(4),* 667-685.

Marr, D. (1982). Vision. San Francisco: Freeman.

Marr, D., & Nishihara, H.K. (1978) Representation and recognition of three dimensional shapes, *Proceeding of the Royal Society of London, B, 200*, 269-291.

Mingolla, E., & Todd, J.T., (1986) Perception of shape from shading, *Biological Cybernetics, 53,* 137-151.

Norman, J.F. & Todd, T.T. (1995) The perception of 3-D structure from contradictory optical patterns, *Perception & Psychophysics, 57*, 826-834.

Norman, J.F., Todd, T.T. & Orban, (2004) Perception of three-dimensional shape from specular highlights, deformations of shading and other types of visual information, *Psychological Science, 15*, 565-570.

Oruç, I., Maloney, L.T. & Landy, M.S. (2003) Weighted linear cue combination with possibly correlated error, *Vision Research, 43*, 2451-2468.

Pasqualotto, A., & Hayward, W.G. (2009) A stereo disadvantage for recognizing rotated familiar object, *Psychonomic Bulletin & Review, 16* , 832-838.

Peissig, J. J., Wasserman, E. A., Young, M. E., & Biederman, I. (2002). Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons. *Vision Research, 42*, 2051–2062.

Pentland, A. (1985). On describing complex surface shapes. *Image and Vision Computing, 3*, 153–162.

Perrett, D.I., Orma, M.W., Harries, M.H. Bevan, R., Hietanen, J.K., Benson, P.J. et Thomas, S. (1991) Viewer-centered and object-centered coding of heads in the macaque temporal cortex, *Experimental Brain Research, 86*, 159-173.

Pizlo, Z. (2008) 3D Shape: Its Unique Place in Visual Perception, MIT Press, Cambridge.

Pizlo, Z., Li, Y., & Francis, G. (2005) A new look at binoculair stereopsis, *Vision Research 45*, 2244-2255.

Pizlo, Z., Li, Y., & Steinman, R.M. (2008) Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose, *Spatial Vision, 21*, 495-508.

Pizlo, Z. & Stevenson, A. (1999) Shape constancy from novel views, *Perception & Psychophysics, 61*, 1299-1307.

Ramachandran, V.S., (1988) Perception of shape from shading, *Nature, 331*, 163-166.

Saunders, J.A. & Backus, B.T. (2006) Perception of surface slant from oriented textures, *Journal of Vision, 6*, 882-897.

Shimojo, S., & Nakajama, Y. (1981) Adaptation to the reversal of binocular depth cues : Effects of wearing left-right reversing spectacles on stereoscopic depth perception, *Perception, 10*, 391-402.

Stevens, K.A., & Brookes, A., (1988) Integrating stereopsis with monocular interpretations of planar surfaces, *Vision Research, 28*, 371-386.

Stevens, K.A., Lees, M., & Brookes, A., (1991) Combining binocular and monocular features, *Perception, 20*, 425-440.

Svarverud, E., Gilson, S.J., Glennerster, A. (2010) Cue combination for 3D location judgements, *Journal of Vision, 10*, 1-13.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geons structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1494–1505.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint?, *Psychological Science, 8*, 282–289.

Tarr, M.J., & Pinker, S. (1989) Mental Rotation and Orientation-Dependence in Shape Recognition, *Cognitive Psychology, 21*, 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame?, *Psychological Science, 1*, 253–256.

Tarr, M. J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape recognition: Further issues, *Psychological Science, 2*, 207–209.

Tarr, M.J., Williams, P., Hayward, W.G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent, *Nature Neuroscience, 1*, 275-277.

Todd, T.T., Norman, J.F., Koenderink, J.J., Kappers, A.M.L. (1997) Effects of texture, illumination and surface reflectance on stereoscopic shape perception, *Perception, 26*, 806-822.

Uka, T., Tanaka, H., Yoshiyama, K., Kato, M.& Fujita, I. (2000) Disparity selectivity of neurons in monkey inferior temporal cortex, *Journal of Neurophysiology, 84*, 120-132.

Ungerleider, L.G., & Mishkin, M. (1982) Two cortical systems, *Analysis of Visual Behavior,* Cambridge, MA, MIT Press, 549-586.

Watt, S.J., Akeley, K., Ernst, M.O, Banks, M.S. (2005) Focus cues affect perceived depth, *Journal of Vision, 5*, 834-862.

Welchman, A.E., Deubelius, A., Conrad, V., Bülthoff, H.H. & Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortx, *Nature Neuroscience, 8,* 820-827.

Wheatsone, C. (1852), Contributions to the physiology of vision part II. On some remarkable, and hitherto unobserved, phaenomena of binocular vision, *Abstract of the papers communicated to the Royal Society of London, 6*.

Yellott, J.I., & Kaiwi, J.L. (1979) Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective, *Perception, 8*, 135-142.

Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z. et Connor, C.E. (2008) A neural code for three-dimensional object shape in macaque inferotemporal cortex, *Nature Neuroscience, 11*, 1352-1360.

Tables

Table1. Linear regression analyses of RTs and ERs as a function of the degree of rotation between shapes.

		Response times			Error rates		
Monocular info	Display mode	slope	Intercept (ms)	R-square	slope	Intercept (p)	R-square
Silhouettes	2D	4.43	689.81	0.94	0.004	0.05	0.94
	Stereo	3.68	681.41	0.92	0.004	0.01	0.99
	Reversed stereo	5.04	681.14	0.97	0.006	0.04	0.96
Shading	2D	4.36	720.7	0.96	0.004	0.02	0.99
	Stereo	2.75	663.77	0.92	0.002	0.01	0.84
	Reversed stereo	3.53	710.97	0.92	0.003	0.02	0.95
Specular highlights	2D	4.84	714.54	0.94	0.006	0.03	0.97
	Stereo	3.77	693.96	0.99	0.004	0.00	0.97
	Reversed stereo	4.09	642.94	0.97	0.004	0.02	0.96
Texture	2D	5.67	676.49	0.98	0.005	0.02	0.99
	Stereo	4.34	730.47	0.95	0.004	0.01	0.98
	Reversed stereo	4.43	702.82	0.98	0.005	0.01	0.97
Specular highlights plus shading	2D	2.91	760.09	0.83	0.004	0.01	0.97
	Stereo	3.55	752.68	0.88	0.003	0.01	0.99
	Reversed stereo	4.88	708.13	0.90	0.005	0.01	0.98
Texture plus shading	2D	3.67	648.48	0.90	0.004	0.02	0.99
	Stereo	3.68	696.15	0.95	0.003	0.03	0.97
	Reversed stereo	4.23	756.99	0.87	0.004	0.03	0.99
Specular highlights plus texture and shading	2D	5.62	704.87	0.98	0.004	0.01	0.94
	Stereo	3.23	710.04	0.95	0.002	0.02	0.96
	Reversed stereo	3.60	775.70	0.94	0.003	0.04	0.97

Table 2. The effects of monocular cues on ERs: the lines indicate conditions showing greater ERs (a) or a greater rotation costs (b) than the conditions indicated by the columns. All the statistics reported are F values. Degrees of freedom in panel (a) are (1, 67), (1, 69), or (1, 70), depending on the particular groups compared. In panel (b), the degrees of freedom are (4, 260), (4, 276), or (4, 280). These variations in degrees of freedom are determined by the fact that some participants had to be eliminated from the data analyses (see above), thereby resulting in unequal group sizes. Pairs of conditions not appearing in the Table show no significant difference between them.

a) Main effect on ERs	Shad/Text/Spec	Shad	Shad/Text
Shad/Spec			
Text	6.38, $p < .001$	6.33, $p < .05$	
Spec	8.58, $p < .005$	8.48, $p < .05$	
Sil	12.85, $p < .001$	12.71, $p < .005$	

b) Interaction with rotation on ERs	Shad/Text/Spec	Shad	Shad/Text
Shad/Spec	4.12, $p < .001$	2.75, $p < .05$	
Text	8.37, $p < .001$	6.62, $p <.001$	4.87, $p < .001$
Spec	8.25, $p < .001$	6.50, $p < .001$	
Sil	9.00, $p < .001$	6.50, $p < .001$	

Figures

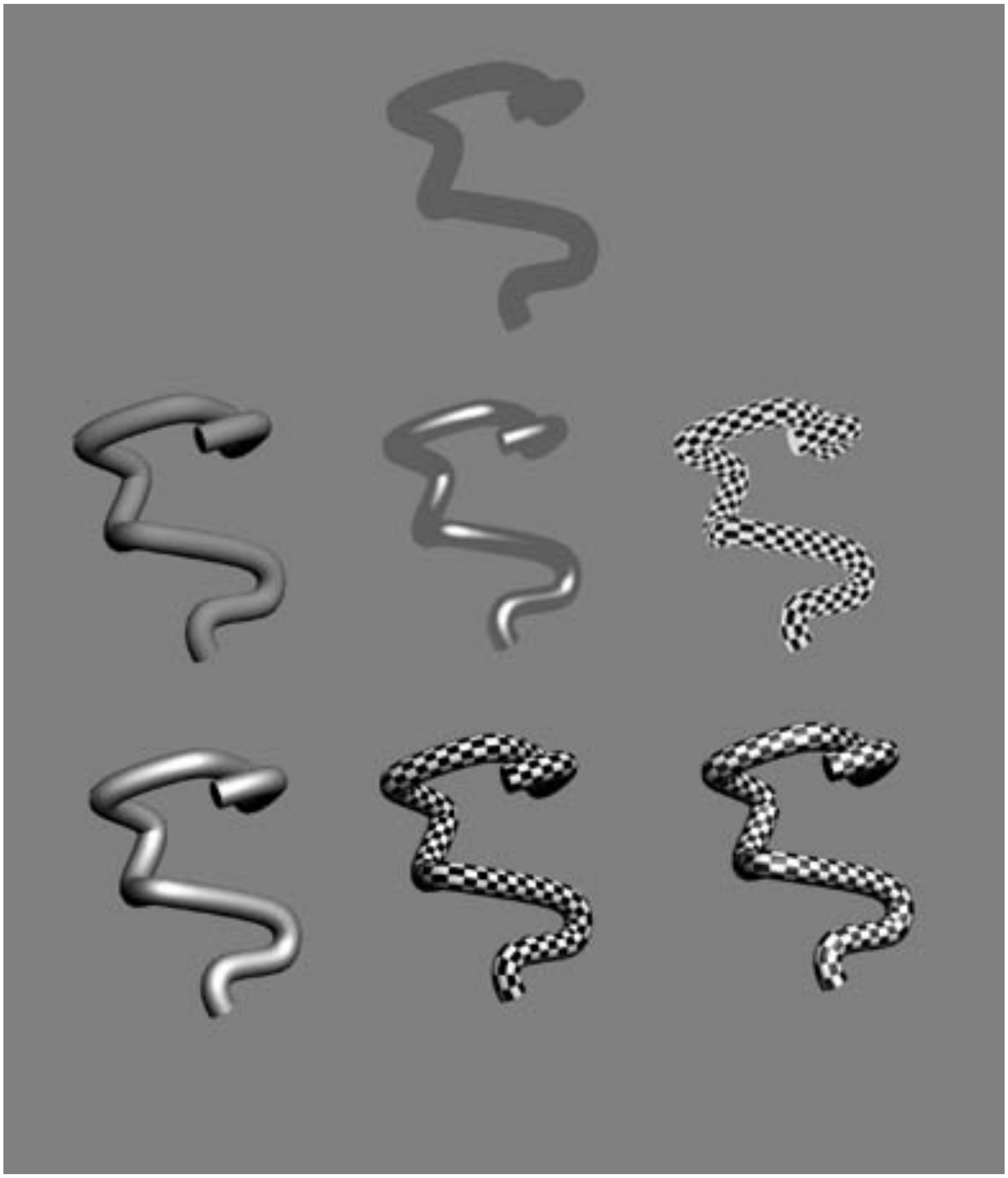


Fig 1.

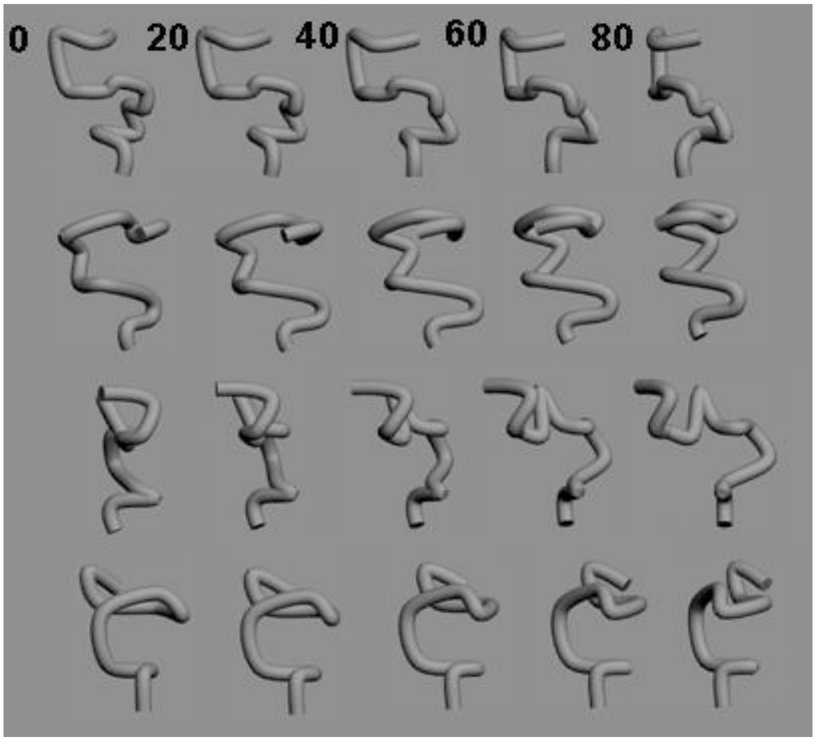
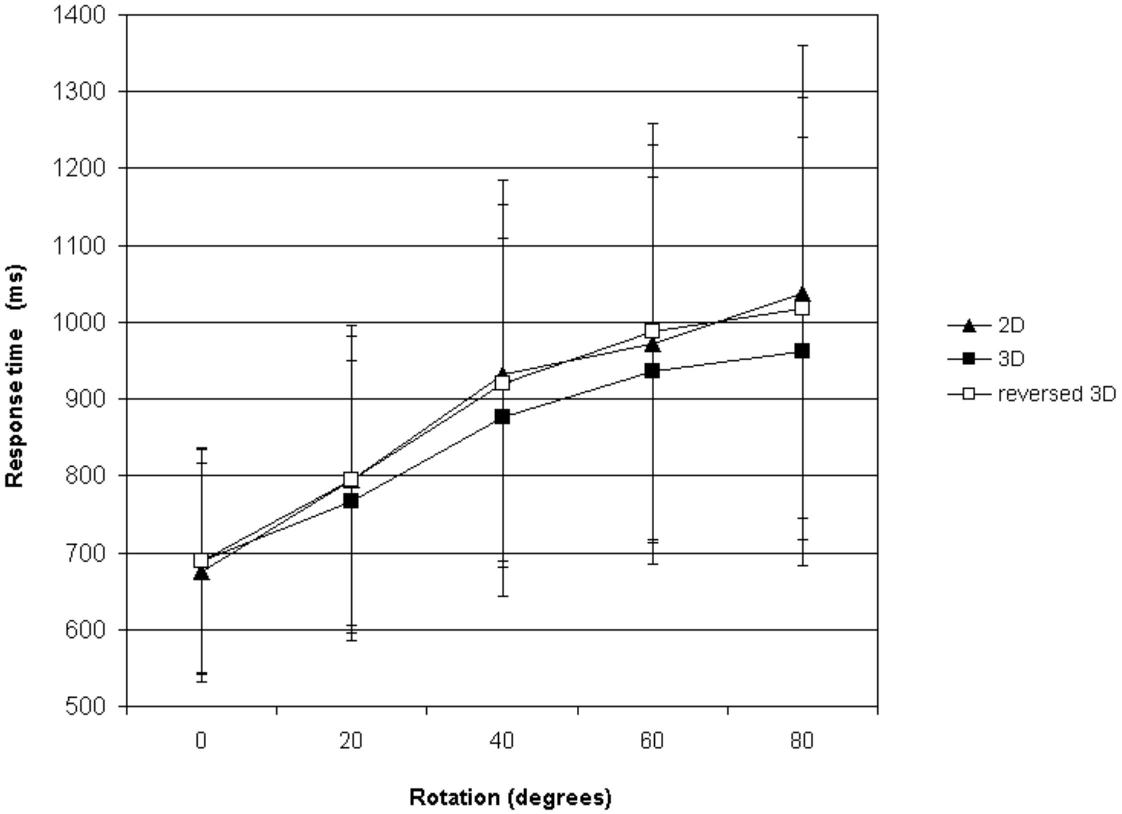
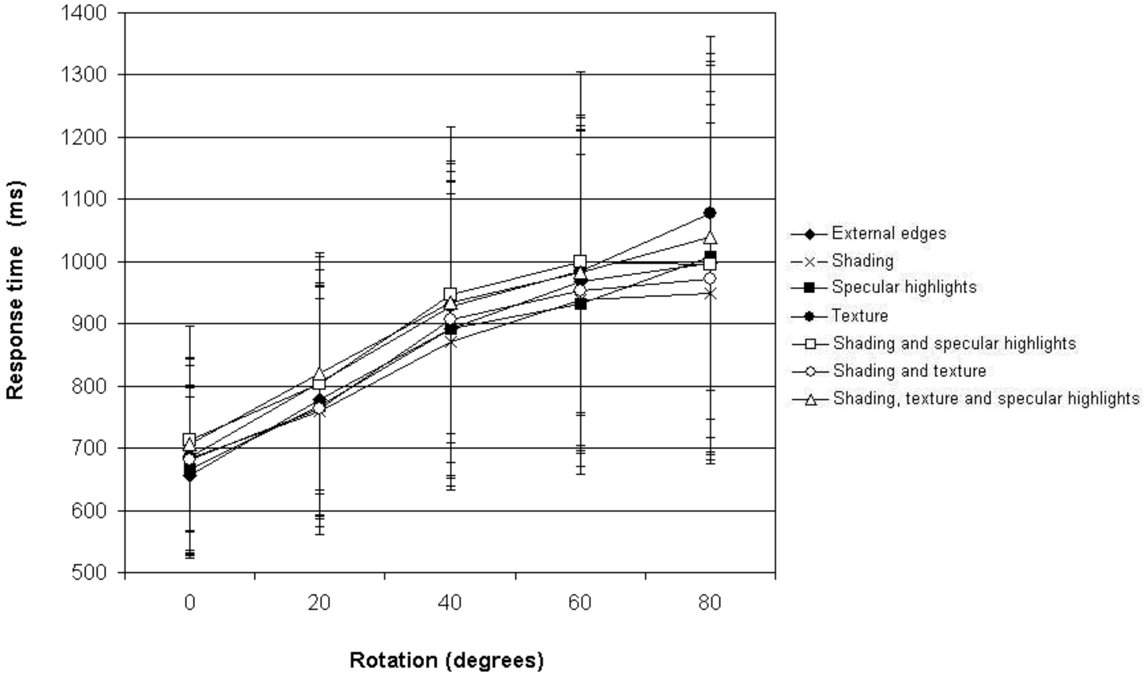


Fig 2.



1400
1300
1200
1100
1000
900
800
700
600
500
Response time (ms)
0
20
40
60
80
Rotation (degrees)
2D
3D
reversed 3D

Fig 3.



1400
1300
1200
1100
1000
900
800
700
600
500
Response time (ms)
0
20
40
60
80
Rotation (degrees)
External edges
Shading
Specular highlights
Texture
Shading and specular highlights
Shading and texture
Shading, texture and specular highlights

Fig 4.

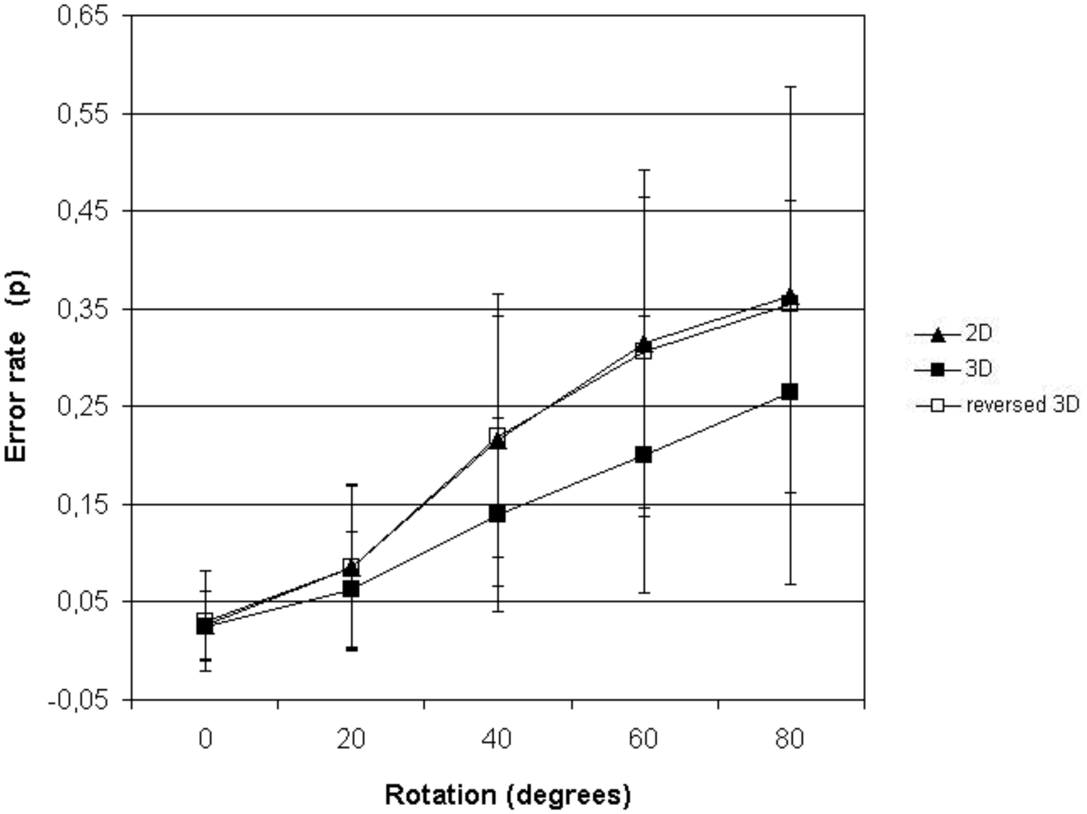

0,65
0,55
0,45
0,35
0,25
0,15
0,05
-0,05
Error rate (p)
0
20
40
60
80
Rotation (degrees)
2D
3D
reversed 3D

Fig 5.

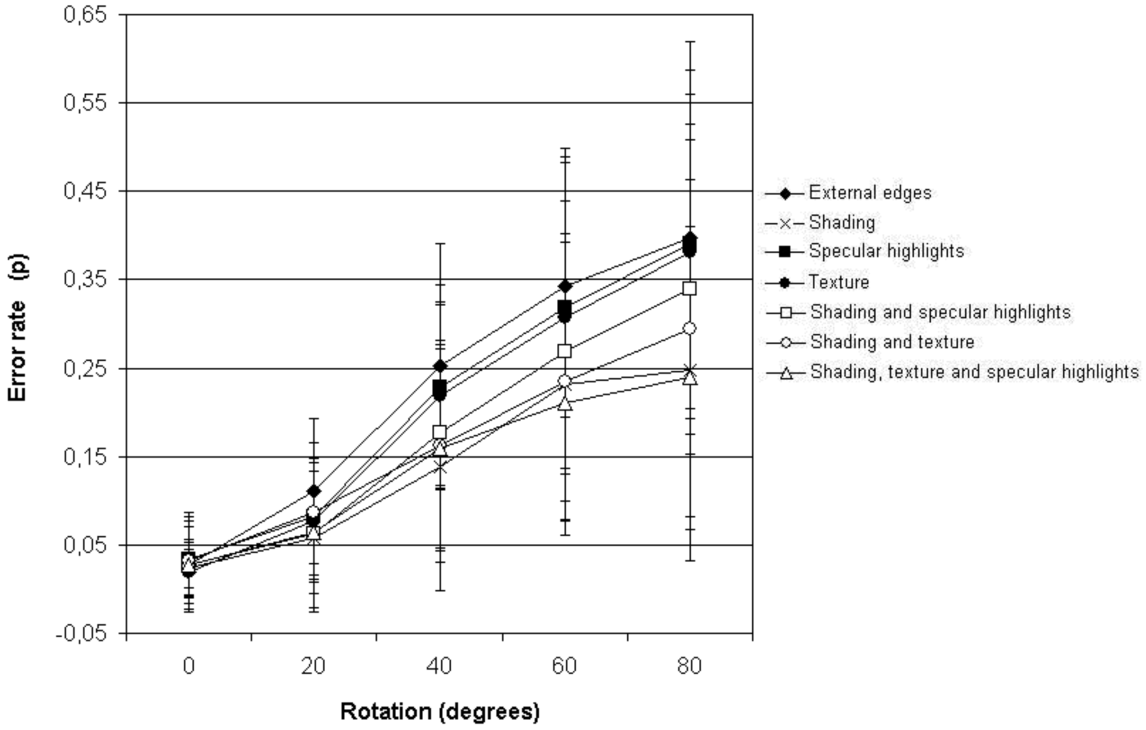

0,65
0,55
0,45
0,35
0,25
0,15
0,05
-0,05
Error rate (p)
0
20
40
60
80
Rotation (degrees)
External edges
Shading
Specular highlights
Texture
Shading and specular highlights
Shading and texture
Shading, texture and specular highlights

Fig. 6