Université de Montréal

Le décours temporel de l'utilisation des fréquences spatiales dans les troubles du spectre autistique

par Laurent Caplette

Département de psychologie, Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de M.Sc. en psychologie

Août 2015

Résumé

Notre système visuel extrait d'ordinaire l'information en basses fréquences spatiales (FS) avant celles en hautes FS. L'information globale extraite tôt peut ainsi activer des hypothèses sur l'identité de l'objet et guider l'extraction d'information plus fine spécifique par la suite. Dans les troubles du spectre autistique (TSA), toutefois, la perception des FS est atypique. De plus, la perception des individus atteints de TSA semble être moins influencée par leurs a priori et connaissances antérieures. Dans l'étude décrite dans le corps de ce mémoire, nous avions pour but de vérifier si l'a priori de traiter l'information des basses aux hautes FS était présent chez les individus atteints de TSA. Nous avons comparé le décours temporel de l'utilisation des FS chez des sujets neurotypiques et atteints de TSA en échantillonnant aléatoirement et exhaustivement l'espace temps x FS. Les sujets neurotypiques extrayaient les basses FS avant les plus hautes: nous avons ainsi pu répliquer le résultat de plusieurs études antérieures, tout en le caractérisant avec plus de précision que jamais auparavant. Les sujets atteints de TSA, quant à eux, extrayaient toutes les FS utiles, basses et hautes, dès le début, indiquant qu'ils ne possédaient pas l'a priori présent chez les neurotypiques. Il semblerait ainsi que les individus atteints de TSA extraient les FS de manière purement ascendante, l'extraction n'étant pas guidée par l'activation d'hypothèses.

Mots-clés: Troubles du spectre autistique; fréquences spatiales; reconnaissance d'objets; décours temporel; échantillonnage; a priori.

Abstract

Our visual system usually samples low spatial frequency (SF) information before higher SF information. The coarse information thereby extracted can activate hypotheses in regard to the object's identity and guide further extraction of specific finer information. In autism spectrum disorder (ASD) however, SF perception is atypical. Moreover, individuals with ASD seem to rely less on their prior knowledge when perceiving objects. In the present study, we aimed to verify if the prior according to which we sample visual information in a coarse-to-fine fashion is existent in ASD. We compared the time course of SF sampling in neurotypical and ASD subjects by randomly and exhaustively sampling the SF x time space. Neurotypicals were found to sample low SFs before higher ones, thereby replicating the finding from many other studies, but characterizing it with much greater precision. ASD subjects were found, for their part, to extract SFs in a more fine-to-coarse fashion, extracting all relevant SFs upon beginning. This indicated that they did not possess a coarse-to-fine prior. Thus, individuals with ASD seem to sample information in a purely bottom-up fashion, without the guidance from hypotheses activated by coarse information.

Keywords: Autism spectrum disorder; spatial frequency; object recognition; time course; sampling; priors.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et des abréviations

ASD: Autism Spectrum Disorder

CGL: Corps genouillé latéral

cpd: Cycles par degré d'angle visuel / Cycles per degree of visual angle

cpi: Cycles par image / Cycles per image

cpo: Cycles par objets / Cycles per object

ERP: Event-related potential

FFT: Fast Fourier Transform

FS: Fréquence spatiale

HSF: High spatial frequency

LSF: Low spatial frequency

PDD-NOS: Pervasive Developmental Disorder Not Otherwise Specified

SF: Spatial frequency

TSA: Troubles du spectre autistique

V1: Cortex visuel primaire

Remerciements

Je tiens à remercier plusieurs personnes qui m'ont soutenu dans l'accomplissement de ce mémoire. Notamment, j'aimerais remercier Nicolas Dupuis-Roy, le doyen de notre lab, sans qui je ne serais pas là où j'en suis. Nicolas, tu as su me prendre sous ton aile alors que je n'étais qu'un jeunot et m'apprendre les rudiments de la méthode en apparence simple mais au fond pas si simple que ça qu'est les Bulles. Merci également de m'avoir laissé une place dans ton projet et dans ton article sur les distances inter-attributs. Merci finalement d'avoir été aussi divertissant. Je me souviendrai (avec attendrissement) de plusieurs moments passés en ta compagnie (e.g., la saga du numéro de téléphone après le Pied de Cochon).

Merci également à Simon Faghel-Soubeyrand (a.k.a. l'undergrad) avec qui j'ai partagé l'émerveillement d'être initié à la fantastique méthode des Bulles. Merci à Jessica T., Jessica R., Alexandre, Solène et Xavier, d'avoir agrémenté mes journées au laboratoire. J'ai trouvé en vous des personnes formidables et vraiment *cool*. Merci aussi à Mercédès, qui venait d'un lab lointain (puis voisin) pour prendre des cafés médiocres (puis fancy) et rire de son rire caractéristique et contagieux. Je remercie également de tout coeur ma famille et mes amis qui m'ont supporté pendant les moments parfois décourageants de ce travail, et surtout qui ont su me faire oublier parfois, pendant de très très brefs moments, mon projet.

Merci également à tous mes prédécesseurs qui ont pavé la voie pour mon projet: je pense plus particulièrement à Verena Willenbockel et à Alan Chauvin. Et un gros merci aux organismes subventionnaires sans qui ce projet n'aurait pas eu lieu: le CRSNG, le CRSH, le FQRNT et Réseau Vision.

Un merci plus particulier à mon co-directeur Bruno Wicker qui a su apporté toute l'expertise en neurosciences sociales et en troubles du spectre autistique nécessaire à l'élaboration de ce projet. Merci de m'avoir fait confiance. Merci énormément également de m'avoir permis d'aller travailler tout un été à Marseille et de m'avoir hébergé. J'ai passé grâce à toi une expérience formidable. Être dehors alors que l'Algérie gagne (ou fait un match nul) pour la première fois, je m'en souviendrai! Merci également de m'avoir fait connaître le bon vin: grâce à toi, je vais être pauvre très rapidement!

Je tiens finalement à remercier mon directeur de recherche, Frédéric Gosselin, sans qui ce mémoire n'aurait pas vu le jour. Frédéric, je te remercie de m'avoir consacré de nombreuses heures et d'avoir été aussi exceptionnellement disponible. Par-dessus tout, je te remercie de m'avoir fait confiance et par le fait même de m'avoir donné confiance en moi et en mes capacités. Je retiendrai plusieurs moments en ta compagnie, notamment les fascinantes discussions intellectuelles, mais aussi avous-le les discussions moins sérieuses et les moments de fous rires avec le lab. À plusieurs autres années encore de collaboration fructueuse!

Introduction

Reconnaitre les objets de notre environnement est une tâche que nous effectuons rapidement et sans effort. Nous y parvenons avec une exactitude remarquable, et ce, malgré le fait que les objets nous apparaissent dans diverses orientations ou conditions d'illumination, ou encore partiellement occlus par d'autres objets. De plus, lorsque nous percevons un certain exemplaire d'objet pour la première fois (e.g., un nouveau modèle de téléphone), nous pouvons dans la très grande majorité des cas le classer dans la catégorie appropriée. Les travaux les plus récents en vision artificielle (e.g., Krizhevsky, Sutskever, & Hinton, 2012) ne permettent toujours pas d'atteindre une exactitude aussi élevée.

Les premiers efforts scientifiques pour expliquer cet aspect important de la perception visuelle se sont appuyés sur des modèles purement ascendants (*bottom-up*). De manière générale, ces modèles postulent que les cellules des aires visuelles primaires détectent certaines propriétés de bas niveau (e.g., traits, contours) du stimulus, puis transmettent cette information aux cellules des aires visuelles de plus haut niveau qui l'intègrent dans le but de la comparer à une représentation stockée en mémoire (e.g., Felleman & van Essen, 1991). Ainsi, l'information serait transmise de manière unidirectionnelle à travers une hiérarchie de régions cérébrales.

Il est impossible, toutefois, que notre système visuel parvienne à une reconnaissance des objets exacte et précise, dans un environnement complexe et changeant, à l'aide de mécanismes de transmission strictement ascendants (e.g., Gilbert & Sigman, 2007; Yuille & Kersten, 2006). Notre cerveau doit également se baser sur des a priori, des connaissances implicites que nous possédons sur le monde, pour donner sens à une information sensorielle ambigüe. De nombreux travaux illustrent d'ailleurs le rôle crucial du traitement descendant (*top-down*) en perception visuelle (e.g., Barceló, Suwazono, & Knight, 2000; Pascual-Leone & Walsh, 2001; Rao & Ballard, 1999; Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999) et soulignent plus spécifiquement l'importante influence de nos a priori sur notre perception (e.g., Kok, Jehee, & de Lange, 2012; Kok, Failing, & de Lange, 2014; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008).

Un a priori fondamental que nous possédons à propos du monde visuel est que celui-ci (ou du moins notre compréhension de celui-ci) est hiérarchique (e.g., Gosselin & Schyns, 2001a, Hegdé, 2008; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), c'est-à-dire que nous concevons les objets comme faisant partie de catégories, qui font elles-mêmes partie de catégories d'un ordre supérieur, etc. Cet a priori a potentiellement un impact important sur la manière de traiter l'information visuelle, puisqu'il s'ensuit logiquement qu'il vaut mieux, dans la plupart des situations naturelles, extraire les propriétés globales d'une image avant ses propriétés plus locales. Pour optimiser nos ressources, il est souvent préférable de commencer par "se renseigner" sur des généralités, puis seulement par la suite de poser des questions plus précises (dépendantes des réponses aux premières questions), plutôt que de tout de suite tenter de deviner l'identité précise de l'objet perçu (voir Hegdé, 2008). Plusieurs études empiriques démontrent que nous extrayons effectivement les propriétés globales des objets et des scènes avant leurs propriétés locales (e.g., Greene & Oliva, 2009; Navon, 1977).

Dans une image, les propriétés locales sont analogues aux hautes fréquences spatiales (FS), tandis que les propriétés globales sont analogues aux basses FS (bien qu'il soit possible de retrouver l'information globale à l'aide des hautes FS également; e.g., Oliva & Schyns, 1997). Nous nous attarderons maintenant aux études portant sur ces FS.

Le décours temporel de l'extraction des fréquences spatiales

Toute image est décomposable, par une opération appelée analyse de Fourier, en un ensemble de grilles sinusoïdales de différentes orientations, fréquences spatiales (FS) et phases. Les basses FS d'une image correspondent aux variations grossières de luminance tandis que les hautes FS représentent les détails et les contours des objets. Les basses FS sont généralement suffisantes pour avoir une ou quelques idées sur l'identité de la catégorie d'objet qui est perçue (Bar, 2003).

Notre système visuel effectue en quelque sorte cette analyse de Fourier. Dès le corps genouillé latéral (CGL), une séparation se fait entre la voie magnocellulaire, dont les neurones sont plus sensibles à de basses FS, et la voie parvocellulaire, dont les neurones sont plus sensibles à de hautes FS (Derrington & Lennie, 1984; Livingstone & Hubel, 1987). Cette spécialisation se poursuit dans le cortex visuel primaire (V1), où les neurones sont organisés en colonnes sensibles à différentes FS (Silverman, Grosof, De Valois, & Elfar, 1989).

Plusieurs études ont démontré que notre système visuel échantillonne l'information de manière *coarse-to-fine*, c'est-à-dire en extrayant les basses FS avant les plus hautes (e.g., Hughes, Nozawa, & Kitterle, 1996; Parker, Lishman, & Hughes, 1992; 1997; Schyns & Oliva, 1994; Watt, 1987). L'extraction précoce des basses FS permettrait au système visuel d'avoir une idée grossière de l'input visuel, afin de générer des hypothèses concernant son identité; ces hypothèses pourraient ensuite guider de manière descendante l'extraction d'information plus détaillée afin de raffiner l'hypothèse initiale (e.g., Bar, 2003; Bullier, 2001; voir également

Friedman, 1979; Grossberg, 1980; Hochstein & Ahissar, 2002; Marr & Poggio, 1979; Ullman, 1984; 1995). Plusieurs études démontrent que les basses FS ou la voie magnocellulaire sont effectivement responsables d'un traitement descendant pendant la reconnaissance d'objets (Bar et al., 2006; Kveraga, Boshyan, & Bar, 2007; Peyrin et al., 2010). Cependant, l'évaluation du décours temporel de l'extraction des FS s'est faite à ce jour de manière relativement imprécise, souvent à l'aide de seulement deux conditions (*coarse-to-fine* vs *fine-to-coarse*, de hautes à basses FS). De plus, les seuils employés dans la définition de "basses" et "hautes" FS sont arbitraires et varient grandement à travers les études (e.g., < 0.4 cycle par degré, cpd, et > 1.4 cpd, dans Boutet, Collin, & Faubert, 2003; < 2.4 cpd et > 8.9 cpd dans Alorda, Serrano-Pedraza, Campos-Bueno, Sierra-Vázquez, & Montoya, 2007; voir exemples dans Caplette, West, Gomot, Gosselin, & Wicker, 2014, en appendice; Willenbockel et al., 2010; Willenbockel, Lepore, Nguyen, Bouthillier, & Gosselin, 2012; voir également discussion dans Hughes et al., 1996). Une compréhension fiable et précise du mécanisme d'extraction des FS est ainsi toujours manquante.

En plus d'être extraites plus tôt, les basses FS semblent également traitées avant les plus hautes FS, la voie magnocellulaire acheminant l'information plus rapidement que la voie parvocellulaire (Bullier & Nowak, 1995; Merigan & Maunsell, 1993; mais voir Skottun, 2015). Un traitement plus rapide des basses FS a été rapporté dans le CGL (Allen & Freeman, 2006) et dans V1 (Mazer, Vinje, McDermott, Schiller, & Gallant, 2002; Purushothaman, Chen, Yampolsky, & Casagrande, 2014).

Extraction vs traitement: note importante

Il y a une importante distinction à faire entre la notion d'échantillonnage (ou d'extraction) et celle de traitement (voir McCabe, Blais, & Gosselin, 2005; VanRullen, 2011). Ces concepts sont souvent utilisés de manière équivalente dans la littérature, ce qui peut mener à une certaine confusion. Lorsqu'une information est captée par le système visuel, celle-ci est transmise à travers une hiérarchie de régions cérébrales. Le traitement désigne les transformations que cette information subit à travers les différentes aires cérébrales dans le but d'accomplir une certaine opération, par exemple la reconnaissance d'un objet. Si une information est traitée plus tôt qu'une autre, cela indique que l'information a pu, pour une raison ou une autre, se rendre du point A au point B plus rapidement.

Lorsqu'on perçoit une scène visuelle, toute l'information n'est pas extraite instantanément: même si on alloue un temps très grand à notre système visuel pour traiter l'information, une trop brève exposition au stimulus ne nous permettra pas de bien identifier toute la scène. Malgré que nous soyons très efficaces pour catégoriser des scènes visuelles avec une exposition de seulement 20 ms (Thorpe, Fize, & Marlot, 1996), notre perception devient plus détaillée seulement si la durée d'exposition est plus longue (Gordon, 2004; Greene & Oliva, 2009). Notre système visuel a en effet une capacité limitée et nous ne pouvons porter attention qu'à une certaine quantité d'information à la fois (voir McCabe et al., 2005). L'échantillonnage désigne cette extraction, par une attention sélective de la part du système visuel, d'une certaine partie de l'information visuelle dans le monde. Ullman (1984) fut le premier à proposer que notre système visuel procède en fait par balayages successifs de l'environnement, extrayant l'information selon une "routine visuelle". Par exemple, lorsque l'on aperçoit un visage, nous ne portons pas attention à tout le visage simultanément, mais plutôt à ses différentes parties dans une séquence particulière (Vinette, Gosselin, & Schyns, 2004). Dans le même ordre d'idées, il est possible que nous extrayions les différentes FS dans une séquence particulière. De manière intéressante, il se peut que les premières étapes de cette routine en informent les suivantes, afin de guider l'extraction et ainsi de la rendre optimale (Ullman, 1984).

Plusieurs auteurs interprètent les résultats d'études portant sur l'*extraction* des FS en s'appuyant sur la notion de *traitement*; pourtant, les deux processus ne sont pas équivalents. Par exemple, plusieurs justifient la perception précoce des basses FS par l'observation que la voie magnocellulaire est plus rapide que la voie parvocellulaire (e.g., Bar, 2003; Peyrin et al., 2010). Cependant, cette interprétation est erronée: même en supposant une vitesse de conduction plus rapide de la voie magnocellulaire, il reste possible que l'information magnocellulaire soit perçue après l'information parvocellulaire, si l'extraction se fait plus tard. Cela semble d'ailleurs être le cas en ce qui concerne le mouvement (surtout traité par la voie magnocellulaire): celui-ci est perçu après la couleur, qui elle, est traitée par la voie parvocellulaire, pourtant plus lente (Bartels & Zeki, 2006). Une explication probable est que l'information de mouvement est extraite après la couleur (voir Dufresne et al., 2013). Une telle extraction tardive pourrait permettre à une information d'arriver à une structure cérébrale donnée à peu près simultanément à une information traitée plus lentement et extraite plus tôt..

Perception visuelle et troubles du spectre autistique

Les troubles du spectre autistique (TSA) sont une famille de troubles neurodéveloppementaux dont les principaux symptômes sont des déficits sociaux et communicationnels. Cependant, ils sont aussi caractérisés par des anomalies sensorielles et perceptuelles, notamment en vision (pour une revue de la littérature, voir Mitchell & Ropar, 2004; Simmons et al., 2009).

Récemment, plusieurs études ont suggéré d'autre part que les individus atteints de TSA sont moins influencés dans leur perception par leurs a priori et leurs connaissances antérieures sur le monde. Par exemple, les autistes reproduisent plus facilement le dessin d'un solide physiquement impossible, étant moins inhibés par l'étrangeté d'une telle figure (Mottron & Belleville, 1993; Mottron, Belleville, & Ménard, 1999). De plus, lorsqu'ils doivent reproduire la projection elliptique d'un cercle, les autistes exagèrent moins la circularité de l'ellipse que les sujets contrôles, étant moins influencés par leur connaissance de la forme réelle (Ropar & Mitchell, 2002). L'hypothèse d'a priori plus incertains (moins précis) dans les TSA a récemment été formalisée dans un cadre Bayésien (Pellicano & Burr, 2012) et proposée en tant qu'explication de plusieurs de leurs symptômes, perceptuels mais aussi sociaux (Lawson, Rees, & Friston, 2013; Sinha et al., 2014; Van de Cruys et al., 2014).

On sait également depuis la description originale de l'autisme par Kanner (1943) que les autistes sont anormalement attirés par les détails d'un objet plutôt que par sa forme globale. De plus, lorsqu'ils doivent extraire l'information locale dans un stimulus hiérarchique, les sujets atteints de TSA ne subissent pas d'interférence de la part de l'information globale, à l'inverse des sujets neurotypiques (Mottron & Belleville, 1993; Rinehart, Bradshaw, Moss, Brereton, & Monge, 2000; Wang, Mottron, Peng, Berthiaume, & Dawson, 2007). Ce style cognitif mettant l'accent sur les détails plutôt que sur le sens global a d'ailleurs été proposé comme explication générale de l'autisme (théorie de la faible cohérence centrale; Frith, 1989; Happé & Frith, 2006).

En lien avec ce focus sur les détails, les individus atteints de TSA semblent avoir une préférence pour les hautes FS. Notamment, ils sont plus sensibles à de hautes FS que des sujets contrôles appariés (tel qu'évalué par leur fonction de sensibilité au contraste; Kéïta, Guy, Berthiaume, Mottron, & Bertone, 2014; mais voir Koh, Milne, & Dobkins, 2010). Également, ils démontrent une composante P1 de plus grande amplitude en réponse à de hautes FS (Vlamings, Jonkman, van Daalen, van der Gaag, & Kemner, 2010). D'autres anormalités relatives au traitement des FS ont également été rapportées: par exemple, ils ne présentent aucune modulation de la N80 par la fréquence spatiale (Jemel, Mimeault, Saint-Amour, Hosein, & Mottron, 2010), contrairement aux neurotypiques.

Peu d'études ont cependant porté sur leur *échantillonnage* des FS; de plus, celles qui l'ont fait n'ont utilisé comme stimuli que des visages. À cet égard, il a été observé que les individus atteints de TSA extraient les hautes FS des visages davantage que les basses FS, alors que les neurotypiques font l'inverse (Deruelle, Rondan, Gepner, & Tardif, 2004; voir aussi Kätsyri, Saalasti, Tiippana, von Wendt, & Sams, 2008). Il reste à savoir si ces résultats s'appliquent à tout objet ou sont restreints à des visages ou à des stimuli sociaux. De plus, l'utilisation des FS n'a été explorée que grossièrement, à l'aide de filtres dont les seuils ont été déterminés arbitrairement. Finalement, nous n'avons pour l'instant aucune indication quant à la dimension temporelle de ce phénomène. Ainsi, nous ne savons pas si l'information est extraite des plus basses aux plus hautes FS chez les individus atteints de TSA; autrement dit, nous ignorons s'ils possèdent un a priori *coarse-to-fine* comme les neurotypiques.

Objectifs et méthode

Dans l'étude principale de ce mémoire, nous vérifierons si les individus atteints de TSA ont un a priori consistant à extraire l'information en basses FS avant celle en hautes FS comparable à celui des neurotypiques. Par le fait même, nous explorerons avec plus de précision que jamais auparavant le décours temporel de l'utilisation des FS chez les neurotypiques.

Pour ce faire, nous emploierons une variante de la méthode *Bubbles* (Gosselin & Schyns, 2001b), qui nous permettra d'étudier l'échantillonnage de stimuli à travers le temps et les FS systématiquement et sans a priori théorique. De manière générale, la méthode *Bubbles* utilise des plages de bruit multiplicatif afin de retrouver l'information utile pour accomplir une certaine tâche (voir Gosselin & Schyns, 2002). Ce bruit peut être appliqué sur n'importe quel espace de représentation des images. Par exemple, la technique a été utilisée sur la dimension temporelle (McCabe et al., 2005; voir aussi Blais, Arguin, & Gosselin, 2013) ainsi que sur la dimension des FS dans l'espace de Fourier (Willenbockel et al., 2010; voir aussi Caplette et al., 2014, en appendice; Tadros, Dupuis-Roy, Fiset, Arguin, & Gosselin, 2013; Thurman & Grossman, 2011; Willenbockel et al., 2012; Willenbockel, Lepore, Bacon, & Gosselin, 2013). Ici, nous combinons pour la première fois ces deux dernières applications et échantillonnons l'espace temps x FS de manière aléatoire (pour une application préliminaire de cette méthode, voir Chauvin et al., 2005).

Présentation des articles et contributions des auteurs

L'article présentant l'étude principale de ce mémoire est rédigé par Laurent Caplette (LC), Bruno Wicker (BW) et Frédéric Gosselin (FG), et est intitulé *Atypical Time Course of*

Object Recognition in Adults with ASD. Le manuscrit est présentement en révision au journal *Psychological Science*. L'idée originale du projet a été élaborée par BW et FG; le programme expérimental a été conçu par LC et FG; les stimuli ont été créés par LC et BW; la passation des participants a été supervisée par LC; le traitement des données et les analyses statistiques ont été accomplies par LC; les résultats ont été interprétés par LC, BW et FG; la recension de la littérature et l'écriture du manuscrit ont été accomplies par LC, BW et FG; FG et BW ont alloué des ressources financières pour la mise en place de ce projet.

L'article *Affective and Contextual Values Modulate Spatial Frequency Use in Object Recognition*, publié dans le journal *Frontiers in Psychology* (2014) et rédigé préliminairement à l'article principal dans le cadre de cette maîtrise, est également inclus dans ce mémoire (Appendice A). Celui-ci consiste en l'application d'une méthode similaire à celle de l'article principal mais avec un échantillonnage des FS seulement (voir Willenbockel et al., 2010), à la reconnaissance d'objets, chez des sujets neurotypiques uniquement. Les auteurs de cet article sont Laurent Caplette (LC), Gregory West (GW), Marie Gomot (MG), Frédéric Gosselin (FG) et Bruno Wicker (BW). L'idée originale du projet a été élaborée par FG et BW; le programme expérimental a été conçu par LC et FG; les stimuli ont été créés par LC, MG, et BW; la passation des participants a été supervisée par LC et GW; le traitement des données et les analyses statistiques ont été accomplies par LC; les résultats ont été interprétés par LC, FG et BW; la recension de la littérature et l'écriture du manuscrit ont été accomplies par LC, FG et BW; GW, FG et BW ont alloué des ressources financières pour la mise en oeuvre de ce projet.

Atypical Time Course of Object Recognition in Adults with ASD

Laurent Caplette[1], Bruno Wicker[2], & Frédéric Gosselin[1]

[1]Département de psychologie, Université de Montréal

[2]Institut de Neurosciences de la Timone, CNRS UMR 7289, Aix-Marseille Université

Abstract

In neurotypical observers, it is widely believed that the visual system samples the world in a coarse-to-fine fashion, whereby a top-down analysis of the fine information is guided by an initial bottom-up analysis of the coarse information. Past studies on ASD have identified atypical responses to fine visual information and a reduced reliance on top-down processes when perceiving objects but did not investigate the time course of the sampling of information at different levels of granularity (i.e. Spatial Frequencies, SF). Here, we examined this question during an object recognition task in ASD and neurotypical observers using a novel experimental paradigm. Our results confirm and characterize with unprecedented precision a coarse-to-fine sampling of SF information in neurotypical observers. In ASD observers, we discovered an inversion of this pattern — a fine-to-coarse sampling of SF information. This suggests that ASD observers rely essentially on a bottom-up extraction of information during object recognition.

Keywords: Autism Spectrum Disorder; Spatial Frequencies; Object Recognition; Coarse-to-Fine; Top-Down.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder whose most prominent symptoms are deficits in social interaction and communication, restricted interests and repetitive behaviors. However, ASD is also characterized by sensory and perceptual peculiarities, as research has increasingly demonstrated in recent years (see Simmons et al., 2009, for a review). There is evidence that individuals with ASD rely more on incoming sensory information and less on top-down mechanisms during perception, when compared with neurotypical individuals (Loth, Gómez, & Happé, 2010; Ropar & Mitchell, 2002). Such a deficit in top-down processing has recently been proposed as a unifying explanation for many of the ASD behavioral symptoms (Pellicano & Burr, 2012; Van de Cruys et al., 2014).

In neurotypical individuals, top-down processing plays a major role in visual object recognition (e.g., Hochstein & Ahissar, 2002). Several authors have submitted that these top-down processes could be initiated by an early extraction of low spatial frequencies (SF). These low SFs would activate an object representation, which in turn would guide the subsequent extraction of higher SFs (e.g., Bar, 2003; Bullier, 2001). In accordance with this proposal, coarse-to-fine SF sampling has been observed in neurotypical subjects (Hughes, Nozawa, & Kitterle, 1996; Schyns & Oliva, 1994), low SFs have been reported to initiate top-down processing during object recognition (Bar et al., 2006; Peyrin et al., 2010), and there is evidence that extraction of SF information is under top-down control (Schyns, Petro, & Smith, 2009; Sowden & Schyns, 2006).

If individuals with ASD rely mostly on bottom-up mechanisms, they might not sample SFs in this coarse-to-fine fashion. Interestingly, many studies indicate atypical SF processing in ASD (see also Van der Hallen, Evers, Brewaeys, Van den Noortgate, & Wagemans, 2015, for a meta-analysis on the related local-global literature in ASD). Unlike neurotypicals, individuals with ASD exhibit a greater P1 amplitude in response to high SFs than to low SFs (Vlamings, Jonkman, van Daalen, van der Gaag, & Kemner, 2010), and a reduced difference between neural responses to high and intermediate SFs (Boeschoten, Kenemans, van Engeland, & Kemner, 2007; Jemel, Mimeault, Saint-Amour, Hosein, & Mottron, 2010). Furthermore, the contrast sensitivity function of individuals with ASD peaks at higher SFs than that of matched neurotypicals (Kéïta, Guy, Berthiaume, Mottron, & Bertone, 2014).

It was shown that individuals with ASD sample more the high SFs than the low SFs of faces, while neurotypicals exhibit the opposite pattern (Deruelle, Rondan, Gepner, & Tardif, 2004). However, these results do not provide any insight whatsoever about the time course of SF extraction. Moreover, most studies that established a coarse-to-fine sampling of SF information in neurotypicals only compared a single low SF condition to a single high SF condition, preventing us to know with precision what are the SFs at play in this process. Here, we used a novel technique based on the Bubbles approach (Gosselin & Schyns, 2001) to map with unprecedented precision the use of SFs across time in both neurotypical and ASD subjects. In essence, we asked subjects, on each trial, to recognize an object from a brief video sampling random SFs on random frames, and reverse correlated the revealed SFs with response accuracy.

Methods

Participants

Fifty-two neurotypical adults and 18 adults with ASD were recruited. A sample size of approximately 20 for the ASD group was targeted at the onset of data collection, as this is usual for a study with this clinical population; for the neurotypical group, a larger sample size of approximately 50 was targeted, because neurotypical participants are easier to recruit and we wanted to increase the overall statistical power of our analyses. ASD participants were diagnosed by an expert psychiatrist or licensed clinical psychologist and the diagnosis had to be recently confirmed, with each having met the criteria for ASD within the past 3 years on the basis of the DSM-IV-TR (American Psychiatric Association, 2000). Brief interviews ensured that none of them suffered from any mental or neurological disorder other than ASD and that they were free of medication. Neurotypical participants were recruited on the campus of the Université de Montréal as a comparison group.

Three neurotypical participants were excluded prior to the analysis: one because he did not complete the first block, one because the quantity of information he required to reach target performance was more than 3 standard deviations over the group mean, and one because his mean response time was more than 3 standard deviations over the group mean. One ASD participant was excluded because the quantity of information he required to reach target performance was more than 3 standard deviations over the group mean. The final ASD group thus included 17 participants (11 males; mean age = 26.88, SD = 8.56), and the final neurotypical group included 49 participants (19 males; mean age = 24.73, SD = 7.94). Subject groups did not differ significantly in age ($t(64) < 1$) or gender ($p = 0.09$, Fisher's exact test).

Participants of both groups had or were completing a post-secondary diploma at the time of the study, and had normal or corrected-to-normal vision. The study was approved by the ethics board of the University of Montreal's Faculty of Arts and Sciences. Written consent from all participants was obtained after the procedure had been fully explained, and a monetary compensation was provided upon completion of the experiment.

Materials

The experimental program ran on Mac Pro (Apple Inc.) computers in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented on Asus VG278H monitors (1920 x 1080 pixels at 120 Hz), calibrated to allow linear manipulation of luminance. Luminance values ranged from 1.6 cd/m^2 to 159 cd/m^2. Chin rests were used to maintain viewing distance at 76 cm.

Stimuli

Eighty-six grayscale object images were selected from the database used in Shenhav, Barrett, & Bar (2013) and from internet searches. Images were 256 x 256 pixels (6 x 6 degrees of visual angle) and median object width was 220 pixels (SD = 47 pixels). The objects were cropped manually and pasted on a uniform mid-gray background. The SF spectrum of every object image was set to the average SF spectrum across all object images using the SHINE toolbox (Willenbockel et al., 2010). This preserved the most important spectral properties of natural objects while eliminating sources of undesired variance in the stimuli. Resulting images had a root mean square (RMS) contrast of 0.20.

On each trial, participants were shown a short video (333 ms) consisting of an object image with random SFs gradually revealed at random time points (e.g., Video S1; Video S2). To create these stimuli, we first randomly generated, on each trial, a matrix of dimensions 256 x 40 (representing respectively SFs from 0.5 to 128 cycles per image or cpi, and frames, each lasting 8.33 ms) in which most elements were zeros and a few were ones. The number of ones was adjusted on a trial-by-trial basis to maintain performance at 75% correct. We then convolved this sparse matrix with a 2D Gaussian kernel (a "bubble"; $\sigma_{SF} = 1.5$ cpi; $\sigma_{time} = 15$ ms). This resulted in the trial's sampling matrix: a SF x time plane with randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic 2D random filters. Finally, these 2D random filters were dot-multiplied by the base image's spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame (see Figure 1 for an illustration of this method). To ensure accurate luminance display, we applied noisy-bit dithering to the final stimuli (Allard & Faubert, 2008).

This random sampling method was preferred to other methods with lower dimensionality because it allowed us to perform a systematic search of the SF x time space and thus to precisely infer which SFs in which time frames led to an accurate response. Moreover, this method is unbiased, as it does not require the selection of a small number of arbitrary cut-offs.

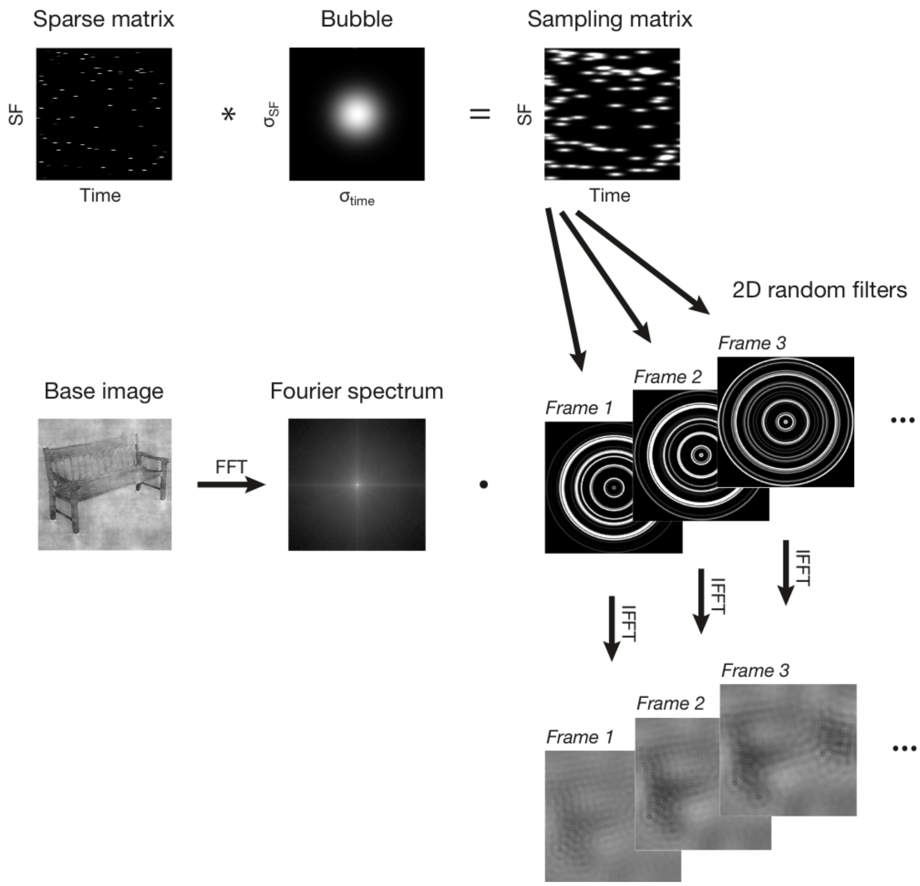

Figure 1. Illustration of the sampling method. On each trial, we randomly generated a matrix of dimensions 256 x 40 (representing respectively SFs and frames) in which most elements were zeros and a few were ones. We then convolved this sparse matrix with a 2D Gaussian kernel (a "bubble"). This resulted in the trial's sampling matrix, shown here as a plane with a number of randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic 2D random filters. Finally, these 2D random filters were dot-multiplied by the base image's spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame.

Procedure

After they had completed a short questionnaire for general information (age, sex, lateralisation, education, language), participants sat in front of a computer monitor, in a dim-lighted room. Participants completed two 500-trial blocks on the first day and two more on another day. Each trial comprised the following consecutive events, on a mid-gray background: a fixation cross (300 ms), a uniform mid-gray field (200 ms), the video stimulus

(333 ms), a fixation cross (300 ms), a uniform mid-gray field (200 ms), and an object name that remained on the screen until a response was provided by the participant or for a maximum of 1 s. Subjects were asked to indicate whether the name matched the object (it did 50% of the time) with a keyboard key press as rapidly and accurately as possible. The number of bubbles (i.e. the quantity of information revealed during the whole video) was adjusted on a trial-by-trial basis using a gradient descent algorithm to maintain performance at 75% correct.

Results

There were no difference between groups in response time (ASD: 708 ms; Controls: 768 ms; $t(64) = 1.07$, $p > .25$) or in the number of bubbles used to perform the task (ASD: 44.80; Controls: 39.00; $t(64) = 1.17$, $p = .25$).

Accuracies and response times were z-scored for each object (to minimize variability due to psycholinguistic factors), for each 500-trial block (to minimize variability due to task learning), and for each subject (to minimize residual individual differences in performance). Trials associated with z-scores over 3 or below -3 (either in accuracy or response time) were discarded (2.27% of all trials). Sampling matrices were also z-scored on each trial; this equalized the importance given to trials irrespective of their number of bubbles.

To uncover which spatial frequencies in which time frames led to accurate object recognition, we performed multiple least-square linear regressions between accuracies and the sparse matrices of the corresponding trials, for each subject. The resulting maps of regression coefficients, or classification images, were convolved with a Gaussian kernel ($\sigma_{SF} = 3.5$ cpi; $\sigma_{time} = 42$ ms) and transformed in z-scores with a bootstrapped sample. We then assessed differences between the groups by performing a random effects analysis on these maps. First, in order to assess information significantly used by each group, we computed a within-group t statistic for every SF-time pixel using all subject classification images from each group; then, in order to assess information used significantly more by one group than by the other, we computed a between-groups t statistic for every SF-time pixel using subject classification images from both groups. Statistical significance of the resulting t maps was then assessed by applying the Pixel test (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005), which corrects for multiple comparisons while taking into account the correlation in the data (see 2D plots in Figure 2). Finally, we computed within-group and between-groups t statistics along time only or along SFs only, and again used the Pixel test to assess statistical significance (see 1D plots in Figure 2).

In neurotypical subjects, SFs between 1 and 21.5 cpi throughout stimulus presentation and SFs between 1 and 36 cpi in the second half of the video ($t(48) > 4.00$, $p < .05$, one-tailed) led to accurate recognition. Averaging over each dimension, the 88-313 ms time window ($t(48) > 2.57$, $p < .05$, one-tailed) and the SFs between .5 and 35 cpi ($t(48) > 3.16$, $p < .05$, one-tailed) were significantly used by neurotypicals (Figure 2). In ASD subjects, three blobs reached statistical significance ($t(16) > 4.93$, $p < .05$, one-tailed): the largest (295 pixels) peaked at 11.5 cpi and 88 ms, the second largest (143 pixels) at 12.5 cpi and 213 ms, and the smallest (44 pixels) at 3.5 cpi and 221 ms. Averaging over each dimension, the 38-279 ms time window ($t(16) > 2.82$, $p < .05$, one-tailed) and the SFs between 1 and 30.5 cpi ($t(16) > 3.59$, $p < .05$, one-tailed; Figure 2) were significantly correlated with accuracy. In the group contrast, the SFs between 23.5 and 26.5 cpi in the 38-80 ms time window led to more accurate recognition in ASD participants than in neurotypical subjects ($t(64) > 4.14$, $p < .05$, two-tailed). Averaging over each dimension, information in the 79-96 ms time window ($t(64) > 2.83$, $p < .05$, two-tailed) and in SFs between 102.5 and 104.5 cpi ($t(64) > 3.36$, $p < .05$, two-tailed) led to more accurate recognition in ASD subjects than in neurotypicals (Figure 2; Figure S1).

To compute an effect size, we then looked at accuracies from specific trials. That is, we compared trials sampling the SF-time space in the most similar way (correlation above 99th percentile) and the least similar way (correlation below 1st percentile) to the in- and out-group *t* maps. The mean accuracy for trials sampling in a similar way to the in-group *t* maps (81.67%) was greater than for trials sampling in a dissimilar way to the in-group *t* maps (66.09%; $t(65) = 10.63$, $p < .001$). It was also greater than for trials sampling in a similar way to the out-group *t* maps (77.86%; $t(65) = 2.82$, $p = .009$). These results confirm that subjects are most accurate when information in the stimulus reflects the information sampling strategy of their own group as uncovered by our regression analyses.

Next, we compared the fit of a coarse-to-fine model with that of a fine-to-coarse model on each subject group *t* map. Specifically, the models were defined by the following inequalities:

$$a_1 + b_1 t < SF < a_2 + b_2 t \ ,$$

where *t* stands for time (s), and a_1, a_2, b_1 and b_2 are free parameters (coarse-to-fine: $b_1, b_2 > 0$; fine-to-coarse: $b_1, b_2 < 0$). The models were fitted to the *t* maps using the Nelder-Mead simplex method. The best coarse-to-fine model fits the neurotypical data better than the fine-to-coarse model, while the opposite is observed for the ASD data (Table 1; Figure 2).

Group	Model	a_1	b_1	a_2	b_2	R^2
Neurotypicals	Coarse-to-fine	0.99	0.00	18.75	55.55	0.81
	Fine-to-coarse	1.47	-0.00	27.44	-0.00	0.75
ASD	Coarse-to-fine	0.84	0.01	28.47	1.91	0.66
	Fine-to-coarse	1.08	-0.00	33.80	-48.92	0.74

Table 1. Parameters (a_1 and a_2 in cpi; b_2 and b_2 in cpi/s) and R^2 of the best-fitting models.

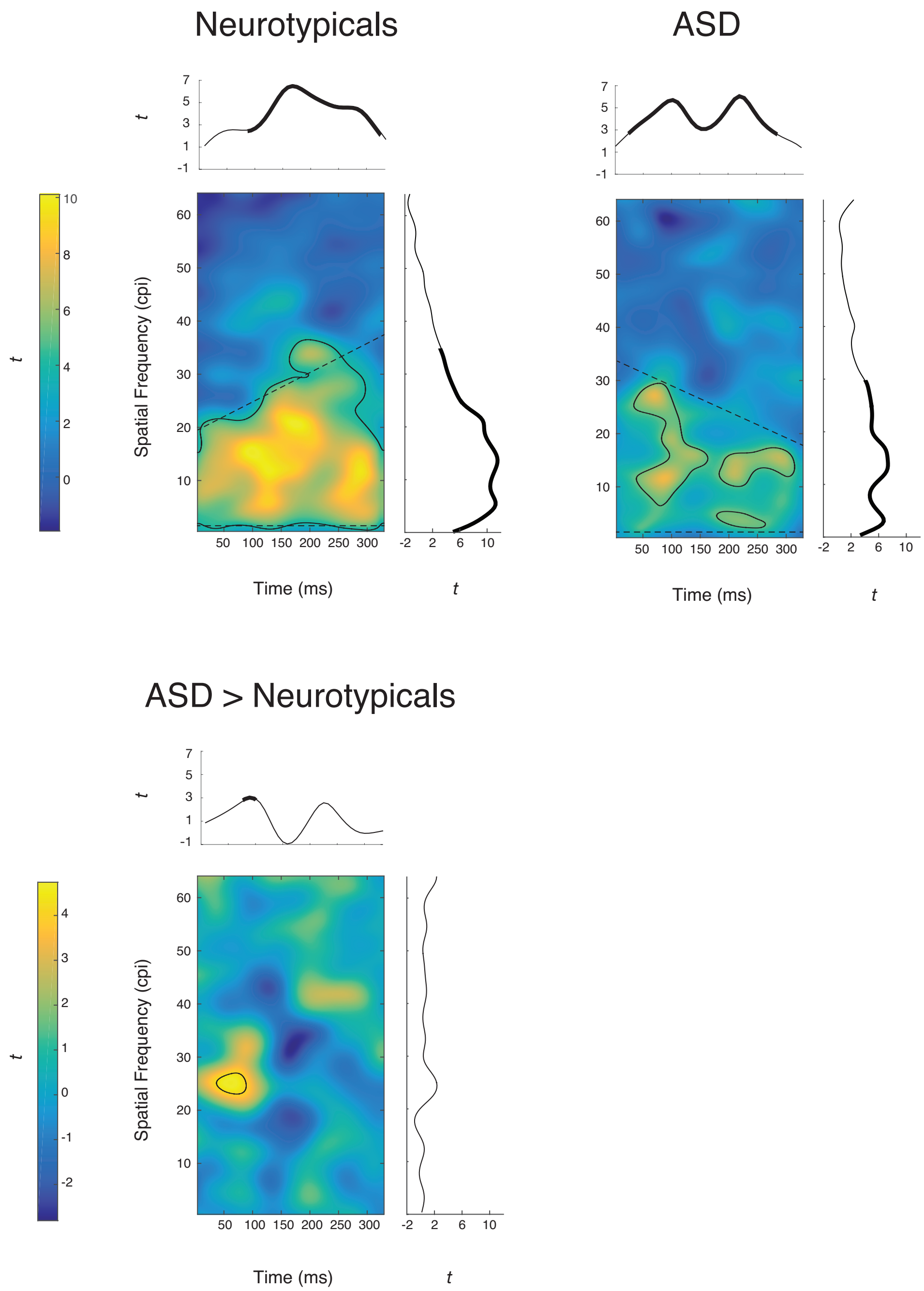

Figure 2. Upper panel: one-sample *t* maps and vectors illustrating how SFs, time frames and SF-time pixels correlate with accurate object recognition, for the neurotypical and ASD groups. Lower panel: two-sample *t* map illustrating the differences in these correlations between both groups. Pixels enclosed by solid lines and bold portions of the vectors are significant ($p < .05$). Dashed lines represent the fitted coarse-to-fine (neurotypicals) and fine-to-coarse (ASD) models (see text for details). Only statistics up to 64 cycles per image (cpi) are shown (see supplementary information for complete *t* maps). Note that color axes in the panels are different.

Discussion

In this study, we investigated with unprecedented precision how neurotypical and ASD subjects sample SFs across time. We first confirmed previous results by revealing that neurotypicals extract SF information in a coarse-to-fine manner. Importantly, our results allowed us to better characterize this mechanism: We discovered that the relationship between the highest SF sampled and time is well described by a line with a slope of about 56 cpi/s and a y-intercept of about 19 cpi, and that the lowest SF sampled is approximately constant at about 1 cpi. This is the first piece of evidence for a hypothesis originally formulated by Ullman (1984; see also Caplette, McCabe, Blais, & Gosselin, in press) according to which low SFs are continuously sampled to activate coarse representations of new objects.

In contrast, our data reveal that ASD observers extract SFs in a fine-to-coarse manner. The relationship between the highest SF the ASD observers sampled and time is well described by a line with a slope of about -49 cpi/s and a y-intercept of about 34 cpi, and the lowest SF they sampled is approximately constant at about 1 cpi. Such fine-to-coarse sampling in ASD subjects might explain the enhanced neural activity they exhibit early in visual processing in response to high SFs (Vlamings et al., 2010). Furthermore, their initial simultaneous sampling of low and high SFs is consistent with a reduced segregation between neural responses to intermediate and high SFs in early processing (Boeschoten et al., 2007; Jemel et al., 2010).

The absence of a coarse-to-fine sampling in individuals with ASD suggests that they rely essentially on a bottom-up extraction of information to categorize objects. Prominent models of object recognition state that the extraction of low SFs activates an object representation that modulates the subsequent extraction of higher SFs in a top-down fashion (Bar, 2003; Bar et al., 2006; Bullier, 2001; Peyrin et al., 2010). In ASD observers, early extraction of low SFs cannot guide the sampling of the diagnostic higher SFs since high and low SFs are sampled equally early soon after stimulus onset. This early unguided sampling of high SFs might also be related with influential theories of perception in ASD positing a bias toward local processing (Happé & Frith, 2006; Mottron et al., 2006).

References

Allard, R., & Faubert, J. (2008). The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution. *Behavior Research Methods*, *40*(3), 735–743. doi:10.3758/BRM.40.3.735

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision). Washington, DC: Author.

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. doi:10.1126/science.8316836

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 449–454. doi:10.1073/pnas.0507062103

Boeschoten, M. A., Kenemans, J. L., Engeland, H. V., & Kemner, C. (2007). Abnormal spatial frequency processing in high-functioning children with pervasive developmental disorder (PDD). *Clinical Neurophysiology*, *118*(9), 2076–2088. doi:10.1016/j.clinph.2007.05.004

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. doi:10.1163/156856897X00357

Bullier, J. (2001). Integrated model of visual processing, *Brain Research Reviews*, *36*, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Caplette, L., McCabe, E., Blais, C., & Gosselin, F. (in press). Categorization of objects, scenes, and faces through time. In C. Lefebvre & H. Cohen (Eds.) *Handbook of categorization in cognitive science* (2nd Edition). Amsterdam: Elsevier.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 659–667. doi:10.1167/5.9.1

Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*(2), 199–210. doi:10.1023/B:JADD.0000022610.09668.4c

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271. doi:10.1016/S0042-6989(01)00097-9

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. doi:10.1007/s10803-005-0039-0

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. doi:10.1016/S0896-6273(02)01091-7

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230. doi:10.1162/jocn.1996.8.3.197

Jemel, B., Mimeault, D., Saint-Amour, D., Hosein, A., & Mottron, L. (2010). VEP contrast sensitivity responses reveal reduced functional segregation of mid and high filters of visual channels in autism. *Journal of Vision*, *10*(6), 13–13. doi:10.1167/10.6.13

Kéïta, L., Guy, J., Berthiaume, C., Mottron, L., & Bertone, A. (2014). An early origin for detailed perception in Autism Spectrum Disorder: biased sensitivity for high-spatial frequency information. *Scientific Reports*, *4*, 5475. doi:10.1038/srep05475

Loth, E., Gómez, J. C., & Happé, F. (2010). When seeing depends on knowing: Adults with Autism Spectrum Conditions show diminished top-down processes in the visual perception of degraded faces but not degraded objects. *Neuropsychologia*, *48*(5), 1227–1236. doi:10.1016/j.neuropsychologia.2009.12.023

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. doi:10.1007/s10803-005-0040-7

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi:10.1163/156856897X00366

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 503–509. doi:10.1016/j.tics.2012.08.009

Ropar, D., & Mitchell, P. (2002). Shape constancy in autism: the role of prior knowledge and perspective cues. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *43*(5), 647–653. doi:10.1111/1469-7610.00053

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dpendent Scene Recognition. *Psychological Science*, *5*(4), 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of Visual Information Integration in the Brain for Categorizing Facial Expressions. *Current Biology, 17,* 1580-

1585.

Shenhav, A., Barrett, L. F., & Bar, M. (2013). Affective value and associative processing share a cortical substrate. *Cognitive, Affective, and Behavioral Neuroscience*, *13*, 46–59. doi:10.3758/s13415-012-0128-4

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. doi:10.1016/j.visres.2009.08.005

Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Neuroscience, 10*(12), 538-545.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. doi:10.1162/jocn.2010.21424

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97-159. doi:10.1016/0010-0277(84)90023-4

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. doi:10.1037/a0037665

Vlamings, P. H. J. M., Jonkman, L. M., van Daalen, E., van der Gaag, R. J., & Kemner, C. (2010). Basic abnormalities in visual processing affect face processing at an early age in autism spectrum disorder. *Biological Psychiatry*, *68*(12), 1107–1113. doi:10.1016/j.biopsych.2010.06.024

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010).

Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. doi:10.3758/BRM.42.3.671

Discussion

De *coarse-to-fine* à *coarse-to-everything*

Dans l'étude décrite dans le corps de ce mémoire, nous avons étudié le décours temporel de l'utilisation des fréquences spatiales (FS) pendant la reconnaissance d'objets, chez des sujets neurotypiques et atteints de TSA. Chez les neurotypiques, la plupart des FS entre 1 et 21.5 cycles par image (cpi) étaient utilisées pendant toute la durée de présentation du stimulus, alors que les FS allant jusqu'à environ 36 cpi étaient utilisées seulement pendant la seconde moitié de la vidéo. Une modélisation a confirmé que le meilleur modèle *coarse-to-fine* (basses à hautes FS), parmi une famille de modèles *coarse-to-fine*, correspondait mieux à ces données que le meilleur modèle *fine-to-coarse* (hautes à basses FS), parmi une famille de modèles *fine-to-coarse*. Une analyse essai par essai a permis de quantifier la taille de l'effet de la présentation d'un tel pattern d'information (15% chez les neurotypiques).

Ces résultats confirment ainsi que le système visuel d'individus neurotypiques extrait les FS selon un mécanisme *coarse-to-fine* et permettent de préciser certains aspects de ce mécanisme qui jusqu'ici demeuraient vagues. Notamment, la taille de l'effet et la forme précise de l'échantillonnage dans l'espace FS-temps demeuraient inconnues: cela était surtout dû à l'utilisation dans les études antérieures d'un faible nombre de conditions (souvent seulement deux sur la dimension des FS: basses vs hautes) et à la grande variabilité des seuils utilisés dans leur définition.

Nos résultats confirment que le système visuel extrait les hautes FS tardivement et démontrent de plus, pour la première fois, qu'il extrait les basses FS en continu (*coarse-to-everything*). Certains auteurs avaient envisagé cette possibilité théoriquement, postulant que notre système visuel doit être sensible à certaines informations en continu afin de détecter de nouveaux objets et d'en créer une représentation de base, laquelle pourrait ensuite guider l'extraction d'information plus spécifique par la suite (McCabe et al., 2005; Ullman, 1984).

Extraction différente dans les troubles du spectre autistique

Chez les individus atteints de troubles du spectre autistique (TSA), la plupart des FS jusqu'à 29 cpi sont utilisées dans les premières 117 ms, alors que seulement des FS inférieures à 17 cpi sont significativement utilisées par la suite, des résultats mieux caractérisés par le meilleur modèle *fine-to-coarse* que par le meilleur modèle *coarse-to-fine*. Une analyse essai par essai a confirmé que la présentation de patterns similaires menait à de meilleures réponses que celle de patterns peu similaires (taille de l'effet = 18% chez les sujets ASD). De plus, en comparant les résultats des groupes neurotypiques et TSA, nous avons observé que des FS autour de 25 cpi (4.17 cpd) entre 38 et 80 ms étaient plus utilisées par les sujets TSA que par les sujets contrôles. Ces FS correspondent en partie aux FS les plus utilisées pour reconnaître les objets, tel qu'évalué avec des images fixes, soit approximativement entre 15 et 26 cpi (Caplette et al., 2014; Appendice A). De plus, ces FS font partie de celles auxquelles les humains sont les plus sensibles (maximum à 3.62 cpd; Watson & Ahumada, 2005).

Ces résultats indiquent que l'a priori qui consiste à traiter l'information des basses aux hautes FS est altéré chez les individus atteints de TSA. Plutôt que d'extraire l'information globale dès le départ afin de guider par la suite l'extraction de hautes FS spécifiques, les sujets atteints de TSA extraient dès le départ toutes les FS utiles à la reconnaissance. Cela ne laisse pas le temps au système visuel d'initier un traitement descendant pendant la reconnaissance de l'objet.

Quelques études antérieures avaient déjà suggéré que l'information globale n'avait pas priorité sur l'information locale dans les TSA, à l'inverse des sujets normaux. Ces études employaient des stimuli hiérarchiques tels que de grandes lettres composées de plus petites lettres (Navon, 1977) afin d'évaluer cet effet. Lorsqu'on leur demandait d'extraire l'information locale uniquement, les individus atteints de TSA n'expérimentaient pas d'interférence de la part de l'information globale, alors que c'est habituellement le cas chez les neurotypiques (Mottron & Belleville, 1993; Rinehart et al., 2000; Wang et al., 2007). Ces résultats peuvent être expliqués par leur échantillonnage atypique des FS: si les individus atteints de TSA échantillonnent rapidement à la fois les basses et les hautes FS, une interférence a peu de chance de survenir, puisque les FS sont toutes traitées sur le même niveau. Nos résultats pourraient également expliquer en partie le focus sur les détails des individus atteints de TSA: en n'orientant pas l'extraction des hautes FS par les basses FS extraites précédemment, ceux-ci doivent possiblement porter une attention beaucoup plus grande aux hautes FS.

De manière intéressante, nous avons également observé que les sujets TSA utilisaient davantage l'information, toutes FS confondues, entre 79 et 96 ms que les neurotypiques, indiquant que le début de l'échantillonnage constituait pour eux une période particulièrement importante. Cela appuie l'hypothèse selon laquelle ils échantillonneraient toute l'information utile dès le départ et que l'échantillonnage subséquent revête peu d'importance pour eux comparativement aux sujets neurotypiques. Finalement, nous avons aussi observé une utilisation plus grande de très hautes FS entre 102.5 et 104.5 cpi de la part des sujets TSA. Cela est cohérent avec certains résultats obtenus dans d'autres études ayant un seuil de hautes FS plus élevé que 25 cpi ou 4.17 cpd (e.g., Deruelle et al., 2004).

Basses vs hautes fréquences spatiales

La définition de ce que constituent des basses et des hautes FS est hautement variable à travers les études (voir exemples cités dans Caplette et al., 2014, en appendice; Willenbockel et al., 2010; 2012). Un avantage important à la méthode que nous avons employé dans notre étude est l'échantillonnage exhaustif de l'espace temps x FS. Ainsi, nous n'avons pas eu à définir des seuils pour nos conditions de basses et de hautes FS. Lors de l'interprétation de nos résultats, nous avons choisi d'employer les termes "basses" et "hautes" de manière relative, en comparant les FS les unes aux autres, plutôt que de manière absolue, en définissant des seuils arbitraires.

Nos résultats suggèrent que les FS autour de 15 (maximalement utilisées par les deux groupes lorsque nous moyennons à travers le temps) et de 25 (utilisation différente entre les groupes) cpi sont particulièrement importantes pour la reconnaissance d'objets. De manière similaire, nous avons observé dans une étude antérieure que des FS entre 15 et 26 cpi étaient maximalement utilisées lors de la reconnaissance d'objets (Caplette et al., 2014; Appendice A). Dans cette étude, nous avions également utilisé une méthode d'échantillonnage aléatoire des fréquences spatiales, mais sans la dimension temporelle, chez des sujets neurotypiques; de plus, l'étude employait en grande partie les mêmes images d'objets. Il est important de noter que beaucoup d'études utilisant des conditions prédéfinies n'échantillonnent pas cette bande de FS souvent considérée comme intermédiaire, décidant plutôt de contraster basses et très hautes FS (e.g., Collin & McMullen, 2005; Harel & Bentin, 2009; Goffaux & Rossion, 2006). Il est évident qu'en faisant cela, ces études manquent malheureusement une partie importante du portrait de la situation. Nous croyons pour cette raison qu'il est important d'échantillonner l'espace de manière exhaustive, ou à tout le moins d'employer plus de deux conditions qui comprennent l'ensemble des FS de l'image.

Conclusion

En premier lieu, notre étude nous a permis de répliquer l'effet *coarse-to-fine* observé dans la plupart des études antérieures; la méthode que nous avons employée nous a également permis de préciser ce mécanisme, en évaluant avec précision les FS concernées et en constatant que les basses FS demeuraient utilisées tout au long de l'échantillonnage. En second lieu, nous avons mis en évidence une altération de ce mécanisme chez les individus atteints de TSA: en effet, les sujets TSA utilisaient davantage de plus hautes FS au début de l'échantillonnage et semblaient extraire les FS selon un pattern *fine-to-coarse*. Cet échantillonnage atypique pourrait partiellement expliquer pourquoi les individus atteints de TSA perçoivent et expérimentent le monde différemment de nous. En conclusion, nos résultats suggèrent que la manière dont l'information est échantillonnée, et non seulement traitée, est importante pour la compréhension de plusieurs déficits visuels associés aux troubles mentaux, et de la cognition humaine en général.

Bibliographie

Allen, E. A., & Freeman, R. D. (2006). Dynamic spatial processing originates in early visual pathways. *Journal of Neuroscience*, *26*(45), 11763–11774. doi:10.1523/JNEUROSCI.3297-06.2006

Alorda, C., Serrano-Pedraza, I., Campos-Bueno, J. J., Sierra-Vázquez, V., & Montoya, P. (2007). Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia*, *45*, 3223–3233. doi:10.1016/j.neuropsychologia.2007.06.017

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. doi:10.1162/089892903321662976

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 449–454. doi:10.1073/pnas.0507062103

Barceló, F., Suwazono, S., & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nature Neuroscience*, *3*(4), 399–403. doi:10.1038/73975

Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research*, *46*(14), 2280–2286. doi:10.1016/j.visres.2005.11.017

Blais, C., Arguin, M., & Gosselin, F. (2013). Human visual processing oscillates: evidence from a classification image technique. *Cognition*, *128*(3), 353–362. doi:10.1016/j.cognition.2013.04.009

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception and Psychophysics*, *65*, 1078–1093. doi:10.3758/BF03194835

Bullier, J. (2001). Integrated model of visual processing, *Brain Research Reviews*, *36*, 96–107. doi:10.1016/S0165-0173(01)00085-6

Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503. doi:10.1016/0959-4388(95)80011-5

Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in Psychology*, *5*:512. doi:10.3389/fpsyg.2014.00512

Chauvin, A., Fiset, D., Ethier, C., Tadros, K., Arguin, M., & Gosselin, F. (2005). Spatial frequency streams in natural scene categorization. Poster session presented at Vision Sciences Society 5th Annual Meeting. *Journal of Vision, 5*(8):603. doi:10.1167/5.8.603

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorisation relies on high spatial frequencies to a greater degree than basic-level categorisation. *Perception and Psychophysics*. *67*, 354–364. doi:10.3758/BF03206498

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*, 219–240. doi:10.1113/jphysiol.1984.sp015498

Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*(2), 199–210. doi:10.1023/B:JADD.0000022610.09668.4c

Dufresne, K., Caplette, L., English, V., Fortin, M., Talbot, M., Fiset, D., et al. (2013). The time course of chromatic and achromatic information extraction in a face-gender discrimination task. Poster session presented at Vision Sciences Society 13th Annual Meeting. *Journal of Vision*, *13*(9):414. doi:10.1167/13.9.414

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. doi: 10.1093/cercor/1.1.1

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*(3), 316–355. doi:10.1037/0096-3445.108.3.316

Frith, U. (1989). Autism: explaining the enigma. Oxford: Blackwell.

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677–696. http://doi.org/10.1016/j.neuron.2007.05.019

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1023–1039. doi:10.1037/0096-1523.32.4.1023

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(4), 760–777. http://doi.org/10.1037/0096-1523.30.4.760

Gosselin, F., & Schyns, P. G. (2001a). Why do we SLIP to the basic level? Computational constraints and their implementation. *Psychological Review*, *108*(4), 735–758. doi:10.1037/0033-295X.108.4.735

Gosselin, F., & Schyns, P. G. (2001b). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. doi:10.1016/S0042-

6989(01)00097-9

Gosselin, F., & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Sciences*, *6*(2), 70–77. doi:10.1016/S1364-6613(00)01838-6

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. http://doi.org/10.1016/j.cogpsych.2008.06.001

Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review*, *87*(1), 1– 51. doi:10.1037/0033-295X.84.5.413

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. doi:10.1007/s10803-005- 0039-0

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1264–1273. doi:10.1037/a0013621

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439. doi:10.1016/j.pneurobio.2007.09.001

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. doi:10.1016/S0896-6273(02)01091-7

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230. doi:10.1162/jocn.1996.8.3.197

Jemel, B., Mimeault, D., Saint-Amour, D., Hosein, A., & Mottron, L. (2010). VEP contrast sensitivity responses reveal reduced functional segregation of mid and high filters of visual channels in autism. *Journal of Vision*, *10*(6), 13–13. doi:10.1167/10.6.13

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*(3), 217–250.

Kätsyri, J., Saalasti, S., Tiippana, K., Wendt, von, L., & Sams, M. (2008). Impaired recognition of facial emotions from low-spatial frequencies in Asperger syndrome. *Neuropsychologia*, *46*(7), 1888–1897. doi:10.1016/j.neuropsychologia.2008.01.005

Kéïta, L., Guy, J., Berthiaume, C., Mottron, L., & Bertone, A. (2014). An early origin for detailed perception in Autism Spectrum Disorder: biased sensitivity for high-spatial frequency information. *Scientific Reports*, *4*, 5475. doi:10.1038/srep05475

Koh, H. C., Milne, E., & Dobkins, K. (2010). Spatial contrast sensitivity in adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(8), 978–987. doi:10.1007/s10803-010-0953-7

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. doi:10.1162/jocn_a_00562

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270. doi:10.1016/j.neuron.2012.04.034

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012) ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing*, *25*, 1106–1114.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232–13240.

doi:10.1523/JNEUROSCI.3481-07.2007

Lawson, R. P., Rees, G., & Friston, K. J. (2013). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*:302. doi:10.3389/fnhum.2014.00302

Livingstone, M., & Hubel, D. (1987). Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception. *Science*, *240*(4853), 740–749. doi:10.1126/science.3283936

Marr D, & Poggio T. (1979) A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci*, *204*(1156), 301–328. doi:10.1098/rspb.1979.0029

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1645–1650. doi:10.1073/pnas.022638499

McCabe, E., Blais, C., & Gosselin, F. (2005). Effective categorization of objects, scenes, and faces through time. In C. Lefebvre & H. Cohen (Eds.) *Handbook of categorization in cognitive science* (pp. 767-791). Amsterdam: Elsevier.

Merigan, W. H., & Maunsell, J. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, *16*, 369–402. doi:10.1146/annurev.ne.16.030193.002101

Mitchell, P., & Ropar, D. (2004). Visuo-spatial abilities in autism: A review. *Infant and Child Development*, *13*(3), 185–198. doi:10.1002/icd.348

Mottron, L., & Belleville, S. (1993). A study of perceptual analysis in a high-level autistic subject with exceptional graphic abilities. *Brain & Cognition*, *23*(2), 279–309. doi:10.1006/brcg.1993.1060

Mottron, L., Belleville, S., & Ménard, E. (1999). Local bias in autistic subjects as evidenced by graphic tasks: Perceptual hierarchization or working memory deficit. *Journal of Child Psychology and Psychiatry*, *40*, 743–755. doi:10.1111/1469-7610.00490

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383. doi:10.1016/0010-0285(77)90012-3

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107. doi:10.1006/cogp.1997.0667

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–160. doi:10.1068/p210147

Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, *26*, 1169–1180. doi:10.1068/p261169

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512. doi:10.1126/science.1057099

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 503–509. doi:10.1016/j.tics.2012.08.009

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. doi:10.1162/jocn.2010.21424

Purushothaman, G., Chen, X., Yampolsky, D., & Casagrande, V. A. (2014). Neural mechanisms of coarse-to-fine discrimination in the visual cortex. *Journal of Neurophysiology*, *112*(11), 2822–2833. doi:10.1152/jn.00612.2013

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Rinehart, N. J., Bradshaw, J. L., Moss, S. A., Brereton, A. V., & Tonge, B. J. (2000). Atypical interference of local detail on global processing in high-functioning autism and Asperger's Disorder. *Journal of Child Psychology and Psychiatry*, *41*(06), 769–778. doi:10.1111/1469-7610.00664

Ropar, D., & Mitchell, P. (2002). Shape constancy in autism: the role of prior knowledge and perspective cues. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *43*(5), 647–653. doi:10.1111/1469-7610.00053

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (2003). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x

Silverman, M. S., Grosof, D. H., De Valois, R. L., & Elfar, S. D. (1989). Spatial-frequency organization in primate striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(2), 711–715. doi:10.1073/pnas.86.2.711

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E.

(2009). Vision in autism spectrum disorders. *Vision Research*, *49*, 2605–2739. doi:10.1016/j.visres.2009.08.005

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., et al. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15220–15225. doi:10.1073/pnas.1416797111

Skottun, B. C. (2015). On the use of spatial frequency to isolate contributions from the magnocellular and parvocellular systems and the dorsal and ventral cortical streams. *Neuroscience & Biobehavioral Reviews*, *56*, 266–275. doi:10.1016/j.neubiorev.2015.07.002

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008) Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. doi:10.1038/nn.2163

Tadros, K., Dupuis-Roy, N., Fiset, D., Arguin, M., & Gosselin, F. (2013). Reading laterally: the cerebral hemispheric use of spatial frequencies in visual word recognition. *Journal of Vision*, *13*(1):4. doi:10.1167/13.1.4

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. doi:10.1038/381520a0

Thurman, S. M., & Grossman, E. D. (2011). Diagnostic spatial frequencies and human efficiency for discriminating actions. *Attention, Perception, & Psychophysics*, *73*(2), 572–580. doi:10.3758/s13414-010-0028-z

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*, 699–703. doi:10.1038/44372

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97-159. doi:10.1016/0010-0277(84)90023-4

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 1–11. doi:10.1093/cercor/5.1.1

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. doi:10.1037/a0037665

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*:365. doi:10.3389/fpsyg.2011.00365

Vinette, C., Gosselin, F., & Schyns, P. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301. doi:10.1016/j.cogsci.2004.01.002

Vlamings, P. H. J. M., Jonkman, L. M., van Daalen, E., van der Gaag, R. J., & Kemner, C. (2010). Basic abnormalities in visual processing affect face processing at an early age in autism spectrum disorder. *Biological Psychiatry*, *68*(12), 1107–1113. doi:10.1016/j.biopsych.2010.06.024

Wang, L., Mottron, L., Berthiaume, C., & Dawson, M. (2007). Local bias and local-to-global interference without global deficit: A robust finding in autism under various conditions of attention, exposure time, and visual angle. *Cognitive Neuropsychology*, *24*(5), 550–574.

doi:10.1080/13546800701417096

Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9), 6–6. doi:10.1167/5.9.6

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, *4*(10), 2006-2021.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 122–135. doi:10.1037/a0016465

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., & Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions — an intracranial ERP study. *Frontiers in Psychology*. *3*:237. doi:10.3389/fpsyg.2012.00237

Willenbockel, V., Lepore, F., Bacon, B. A., & Gosselin, F. (2013). The informational correlates of conscious and nonconscious face-gender perception. *Journal of Vision*, *13*(2):10. doi:10.1167/13.2.10

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. doi:10.1016/j.tics.2006.05.002

Appendice A: Article 2

Affective and contextual values modulate spatial frequency use in object recognition

Laurent Caplette[1], Greg L. West[1], Marie Gomot[2], Frédéric Gosselin[1] & Bruno Wicker[3]

[1]CERNEC, Département de psychologie, Université de Montréal

[2]UMR-S ‘Imaging and Brain’, INSERM U930, CNRS ERL3106, Université François-Rabelais de Tours

[3]Institut de Neurosciences de la Timone, CNRS UMR 7289, Aix-Marseille Université

Abstract

Visual object recognition is of fundamental importance in our everyday interaction with the environment. Recent models of visual perception emphasize the role of top-down predictions facilitating object recognition via initial guesses that limit the number of object representations that need to be considered. Several results suggest that this rapid and efficient object processing relies on the early extraction and processing of low spatial frequencies (SF). The present study aimed to investigate the SF content of visual object representations and its modulation by contextual and affective values of the perceived object during a picture-name verification task. Stimuli consisted of pictures of objects equalized in SF content and categorised as having low or high affective and contextual values. To access the SF content of stored visual representations of objects, SFs of each image were then randomly sampled on a trial-by-trial basis. Results reveal that intermediate SFs between 14 and 24 cycles per object (2.3 to 4 cycles per degree) are correlated with fast and accurate identification for all categories of objects. Moreover, there was a significant interaction between affective and contextual values over the SFs correlating with fast recognition. These results suggest that affective and contextual values of a visual object modulate the SF content of its internal representation, thus highlighting the flexibility of the visual recognition system.

Keywords: object recognition, internal representations, affective value, context, spatial frequencies.

Introduction

Rapid and accurate visual recognition of everyday objects encountered in different orientations, seen under various illumination conditions, and partially occluded by other objects in a visually cluttered environment is necessary for our survival. The first theoretical efforts to explain this feat relied on purely bottom-up mechanisms in the visual system: cells in early visual areas would be sensitive to low-level features and cells in higher areas would integrate this information in order to then match it to a representation in memory (e.g., Maunsell & Newsome, 1987). However, it is improbable that feedforward pathways alone can account for object recognition because of their severely limited information processing capabilities (Gilbert & Sigman, 2007). Moreover, since these early theoretical efforts, the essential role of such feedback mechanisms in vision has been amply demonstrated (e.g., Barceló, Suwazono, & Knight, 2000; Pascual-Leone & Walsh, 2001; Rao & Ballard, 1999; Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999). Nowadays, most top-down models of object recognition (e.g., Friston, 2003; Grossberg, 1980; Ullman, 1995) propose that the search for correspondence between the input pattern and the stored representations is a bidirectional process where the input activates bottom-up as well as top-down streams that simultaneously explore many alternatives; object recognition is achieved when the counter streams meet and a match is found. The content of these stored representations could depend on several factors such as task requirements (e.g., perception or action, basic-level vs. superordinate-level categorisation) or categorical properties of the object (e.g., animate vs. inanimate, affective vs. non affective, social vs. non social; Logothetis & Scheinberg, 1996). Understanding the properties of the stored representations that lead to the generation of predictions thus is an important unexplored issue. In particular, it remains to be understood if different representational systems are used during recognition of different categories of visual objects.

Building on the predictive account of visual object recognition, Bar (2003) proposed a brain mechanism for the cortical activation of top-down processing during object recognition, where low spatial frequencies (LSFs) of the image input are projected rapidly and directly through quick feedforward connections, from early visual areas into the dorsal visual stream. Such LSF information activates a relatively small set of probable candidate interpretations of the visual input in higher prefrontal integrative centres. These initial guesses are then back-projected along the reverse hierarchy to guide further processing and gradually encompass high spatial frequencies (HSFs) available at lower cortical visual areas. This proposal is supported by neurophysiological, computational and psychophysical evidence that LSFs are processed earlier than HSFs (Bredfeldt & Ringach, 2002; Mermillod, Guyader, & Chauvin, 2005; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Schyns & Oliva, 1994; Watt, 1987; for reviews, see Bar, 2003; Bullier, 2001; Hegdé, 2008) and that top-down processing in visual recognition relies on LSFs (Bar et al., 2006); moreover, magnocellular projections, which are more sensitive to LSFs (Derrington & Lennie, 1984), seem to be implicated in initiation of top-down processing (Kveraga, Boshyan, & Bar, 2007). Stored internal representations may thus be biased toward LSFs, since objects would be primarily matched in memory with an LSF draft.

Only a handful of studies have focused on the effect of specific SF band filtering during object recognition. In a name-picture verification task, low-pass filtering selectively impaired subordinate-level category verification (e.g., verify the "Siamese" category instead of the “animal” category at the superordinate level or the “cat” category at the basic level), while having little to no effect on basic-level category verification, suggesting that basic-level categorisation does not particularly rely on low spatial frequencies (Collin & McMullen, 2005). On the other hand, Harel and Bentin (2009) reported that subordinate-level categorisation was impaired by the removal of HSFs, but also that basic-level categorisation was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially useful for recognition at the basic level. Finally, using a superordinate-level categorisation task, Calderone et al. (2013) reported no difference in accuracy or response times between LSFs and HSFs. Overall, these studies suggest that, although this seems a bit different for subordinate-level categorisation, neither LSFs or HSFs have a privileged role in object recognition. Even if LSFs do initiate a top-down processing, this suggests that their overall role in recognition is negligible; other SFs (neither low or high), however, may have a preponderant role.

Intrinsic properties of visual objects such as their affective value or contextual associativity may modulate the content of internal representations. Because of their great adaptive value, emotional objects might necessitate fast recognition, to facilitate an immediate behavioural response; this is likely to apply to both dangerous and pleasant stimuli, the former threatening survival and the latter promoting it (Bradley, 2009). In fact, the brain’s prediction about the identity of a visual object may be partly based on its affective value, i.e. prior experiences of how perception of a given object has influenced internal body sensations. As such, affective value could be not just a label or judgment applied to the object post-recognition, but rather an integral component of mental object representations (Lebrecht, Bar, Barrett & Tarr, 2012) and could act as an additional clue to the object's identity to facilitate its recognition (Barrett & Bar, 2009). Since emotional objects need to be processed quickly, it is likely that LSFs, which are extracted rapidly, are particularly important for their recognition. In agreement with this idea, there is some evidence that LSFs are more present in representations of objects with strong affective value than in representations of neutral objects. Mermillod, Droit-Volet, Devaux, Schaefer and Vermeulen (2010) reported that threatening stimuli were recognized faster and more accurately than neutral ones with LSFs but not with HSFs. Other behavioural and neuroimaging studies also suggested an interaction between emotional content and LSFs in various perceptual tasks. Bocanegra and Zeelenberg (2009), for instance, observed that in a Gabor orientation discrimination task, briefly presented fearful faces improved subjects' performance with LSF gratings while impairing it with HSF gratings. Moreover, early ERP amplitudes sensitive to affective content were found to be greater when unpleasant scenes were presented intact or in LSFs rather than in HSFs (Alorda, Serrano-Pedraza, Campos-Bueno, Sierra-Vázquez & Montoya, 2007). In the same vein, Vuilleumier, Armony, Driver and Dolan (2003) observed that the amygdala responded to fearful faces only if LSFs were present in the stimulus. In an intracranial ERP study where subjects were presented with both visible and invisible (masked) faces, Willenbockel, Lepore, Nguyen, Bouthillier and Gosselin (2012) found that amygdala activation correlated mostly with SFs around 2 and 6 cycles/face, while insula activation correlated mostly with slightly higher SFs near 9 cycles/face. All these results suggest that the internal representations of objects with affective value would comprise more LSFs than representations of neutral objects.

Relatedly, the contextual associativity of a visual object — “what other objects or context might go with this object?” (Bar, 2004; Fenske, Aminoff, Gronau & Bar, 2006) — could also impact on the SF content of its mental representation. It has been shown that recognition of an object that is highly associated with a certain context facilitates the recognition of other objects that share the same context (e.g., Bar & Ullman, 1996). A lifetime of visual experience would lead to contextual associations that guide expectations and aid subsequent recognition of associated visual objects through rapid sensitization of their internal representations (Biederman, 1972, 1981; Palmer, 1975; Biederman, Mezzanotte & Rabinowitz, 1982; Bar & Ullman, 1996). This associative processing is quickly triggered merely by looking at an object and would be critical for visual recognition and prediction (Aminoff et al., 2007; Bar and Aminoff, 2003). It has been suggested that the rapidly extracted LSFs of an object image are sufficient to activate these associated representations, and thus that the representations of contextual objects are likely to be biased toward LSFs (Bar, 2004, Fenske et al., 2006). However, this hypothesis has never been tested directly.

Affective and contextual values may also interact, so that representations of visual objects with affective value could be modulated by their contextual value or vice-versa (e.g., Brunyé et al., 2013; Shenhav et al., 2013; Storbeck & Clore, 2005). Indeed, the affective value of a given object is often defined by the context to which it has been associated to in memory. For example, a tomb elicits sadness, not because it is inherently sad, but because it evokes a context of cemetery/death. As such, affective objects might be differentially represented whether or not their affective value originates from their associated contexts. Interactions between both psychological properties have been reported. For instance, our affective state influences the breadth of the associations we make (Storbeck & Clore, 2005) and conversely, the generation of associations influences our affective state (Brunyé et al., 2013). Also, it seems that associative and affective processing both take place in the medial orbitofrontal cortex, and that both contextual and affective values might in fact relate to a more unified purpose (Shenhav et al., 2013).

The current study examined the SF content of stored internal representations of visual objects with different affective and contextual values, by evaluating what are the SFs in the stimuli that correlate with fast and accurate identification. Stimuli consisted of pictures of objects equalized in SF content and categorised as having low or high affective and contextual values. The SFs of these stimuli were randomly sampled on a trial-by-trial basis while subjects categorized the objects portrayed in the images. By varying affective value, contextual value and spatial frequencies available in the object image altogether, we aimed to clarify their roles in visual recognition, and to study potential interactions between them.

Methods

Participants

Forty-seven healthy participants (33 males) with normal or corrected-to-normal visual acuity were recruited on the campus of the Université de Montréal for an object recognition study. Participants were aged between 19 and 31 years (M = 23.04; SD = 3.13) and did not suffer from any reading disability. A written informed consent was obtained prior to the experiment, and a monetary compensation was provided upon its completion.

Apparatus

The experimental program was run on a Mac Pro computer in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A refresh rate of 120 Hz and a resolution of 1920 x 1080 pixels were set on the Asus VG278H monitor used for stimuli presentation. The relationship between RGB values and luminance levels was linearized. Luminance depth was 8 bits, and minimum and maximum luminance values were 1.1 cd/m^2 and 134.0 cd/m^2, respectively. A chin rest was used to maintain viewing distance at 76 cm.

Stimuli

Selection and validation. One hundred fifty six object images were pre-selected mainly from the database used in Shenhav et al. (2013) but also from Internet searches. Each object image was presented to 30 raters who decided either i) if they associated the object to a particular emotion, and if so, to which one or ii) if they associated the object to a particular context, and if so, to which one. For the experiment, we selected 18 objects with clear consensus (or absence of) regarding their contextual and affective values in each of our four object categories: contextual emotional, non-contextual emotional, contextual neutral and non-contextual neutral (figure 1, table S1). Clear consensus about high affective or high contextual value meant that an object was associated to the same context or to the same emotion by more than 75% of raters; and clear consensus about low affective or contextual value meant that an object was associated to no particular context or emotion by more than 75% of the raters. Fifty-one of the selected images came from the Shenhav et al. (2013) database, and our affective and contextual ratings for these images closely matched theirs.

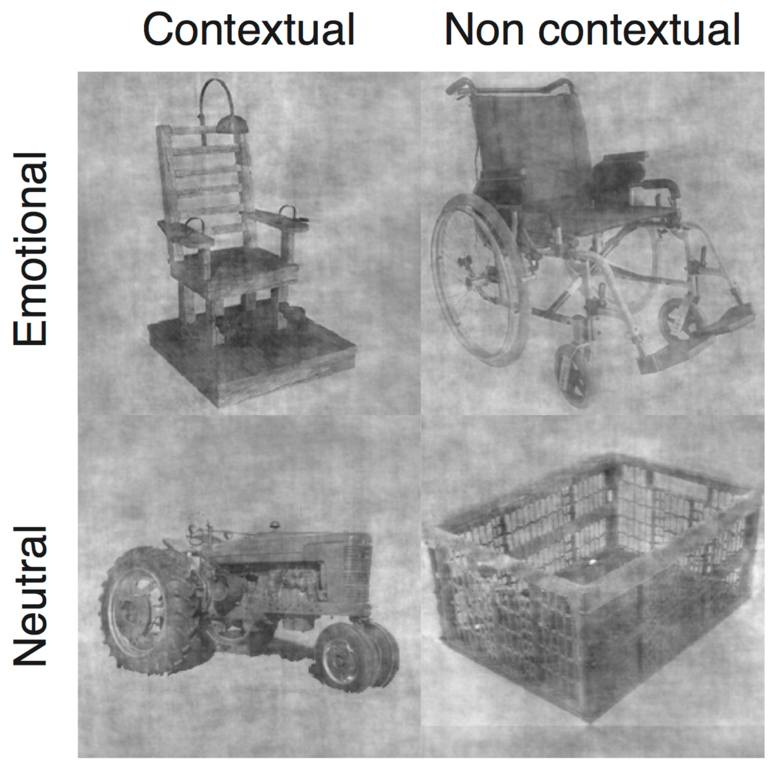

Figure 1. Example images for each of the four categories of objects.

Control of low-level features. Stimuli thus consisted of 72 grayscaled object images of 256 x 256 pixels presented on a mid-gray background. The images subtended 6 x 6 degrees of visual angle. Median object width was equal to 237 pixels. To target our investigation on stored internal representations and get rid of a potential interaction between the visual input and the representation, spatial frequency content and luminance were equalized across stimuli using the SHINE toolbox (Willenbockel et al., 2010a). Resulting images had a RMS contrast of 0.075. We reduced the undesired impact of psycho-linguistic factors, such as word length and lexical frequency, on response times by transforming these into z-scores for every object.

For example, we computed the mean and standard deviation of the RTs of the correct positive trials in which the electric chair was presented, and we used these statistics to transform those RTs into z-scores. We did the same for all the other objects. As a result, the means and standard deviations of the RTs associated with every word were strictly identical, and all RT variations due to differences between the words were eliminated.

Sampling. SF content of the images properly padded was extracted via Fast Fourier Transform (FFT) and randomly filtered at each trial, according to the SF Bubbles method (Willenbockel et al., 2010b). In short, each spatial frequency filter was created by first generating a random vector of 10,240 elements consisting of 20 ones (the number of bubbles) among zeros. Second, the resulting vector was convolved with a Gaussian kernel that had a standard deviation of 1.8. Third, the vector was log transformed so that the SF sampling approximately fit the SF sensitivity of the human visual system (see De Valois & De Valois, 1990). The resulting sampling vector contained 256 elements representing each spatial frequency from 0.5 to 128 cycles per image. To create the two-dimensional spatial frequency filtered images, vectors were rotated about their origins and dot-multiplied with the FFT amplitudes (see Willenbockel et al., 2010b, for methodological details). Thus, several SF bandwidths were revealed in each stimulus; and objects were presented several times with different SF bandwidths revealed every time (figure 2).

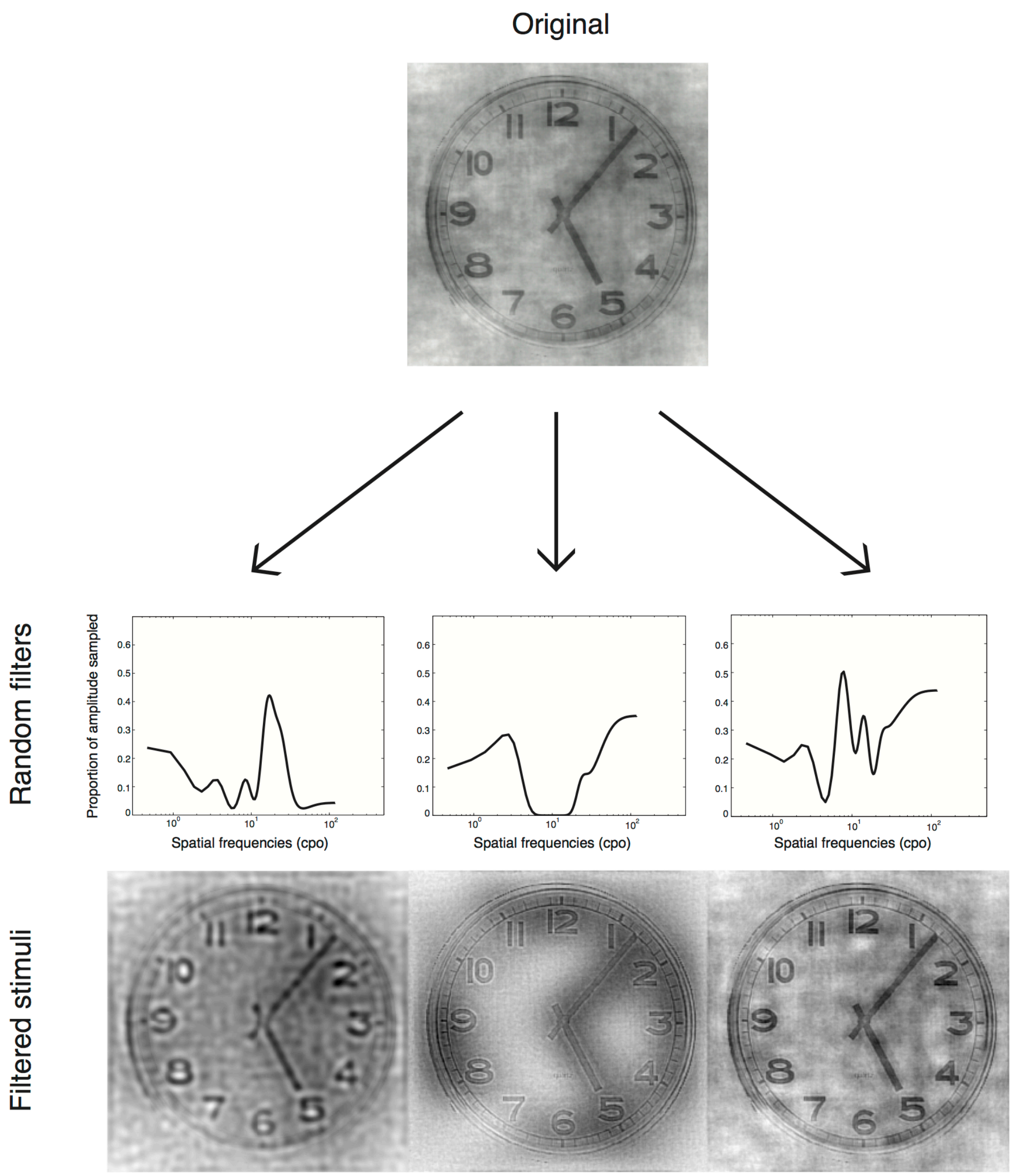

Figure 2. Examples of stimuli presented in the experiment. These are generated by applying random filters to a base image.

Procedure

After they had completed a short questionnaire for general information (age, sex, education, language, etc.), participants sat comfortably in front of a computer monitor, in a dim-lighted room. Participants did two 500-trial blocks, with a short break in between. Each trial began with a central fixation cross lasting 300 ms, followed by a blank screen for 100 ms, the SF-filtered random object image for 300 ms, a central fixation cross for 300 ms, a blank screen for 100 ms, and finally a matching or mismatching object name that remained on the screen until the participant had answered or for a maximum of 1000 ms. Subjects were asked to indicate with a keyboard key press as accurately and rapidly as possible whether or not the name matched the object depicted in the image. This picture-name verification task was chosen because it imposes a specific level of categorisation to subjects (we chose the basic-level) without focusing attention explicitly on either affective or contextual value of the object. Name and object matched on half the trials.

Spatial frequency data analysis

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between RTs and corresponding sampling vectors. Only correct positive trials (i.e., when the name matched the object, and the participant answered correctly) were included in the analysis. RTs were first z-scored for every object to minimize undesired sources of variability pertaining to psycho-linguistic factors such as word length and lexical frequency (see Stimuli: Control of low-level features). They were further z-scored for each condition in each subject's session to diminish variability due to task learning. Trials associated with z-scores over 3 or below -3 were discarded (< 1.8% of trials).

We call the resulting vectors of regression coefficients classification vectors. We first contrasted the classification vector for all objects against zero to examine what were the spatial frequencies used in general, regardless of affective or contextual values. We then contrasted the classification vectors for all emotional objects and all neutral objects, and the ones for all contextual objects and all non-contextual objects, to assess the main effects of contextual and affective values. Next, we examined if there was an interaction between these two dimensions. To do so, we contrasted classification vectors of all four subcategories of objects by applying the following formula:

$$(A_1B_1 - A_1B_2) - (A_2B_1 - A_2B_2),$$

where A represents emotional value, B represents contextual value, and the number represents the level of the variable. We finally investigated the simple effects by comparing the conditions pairwise. The statistical significance of the resulting classification vectors was assessed by applying the Cluster test (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). Given an arbitrary z-score threshold, this test gives a cluster size above which the specified p-value is satisfied. We used this test rather than the Pixel test (Chauvin et al., 2005) because it is in general more sensitive, allowing us to detect weaker but more diffuse signals. Here, we used a threshold of ±3 ($p<0.05$, two-tailed). We report the size k of the significant cluster and its maximum Z-score Z_{max}. We implemented the Cluster tests as bootstraps (Efron & Tibshirani, 1993); that is, we repeated all regressions 10,000 times pairing the sampling vectors with transformed RTs randomly selected in the observed transformed RT distribution. This resulted in 10,000 random classification vectors per condition. We used these random classification vectors to transform the elements of the observed classification vectors into z-scores and estimate their p-values. We corrected p-values for multiple comparisons in the pairwise comparisons by implementing Hochberg's step-up procedure (Hochberg, 1988).

Results

Effects of condition and spatial frequencies on accuracy

The mean accuracy was 87.49% (SD = 7.63). To analyse possible effects of condition on accuracy, without taking SFs into account, we first conducted a 2 (Context: non-contextual or contextual) x 2 (Emotion: neutral or emotional) repeated-measures ANOVA on mean accuracies per participant. There was an effect of contextual value [$F(1,46) = 39.83$, $p<0.001$, $\eta_p^2 = 0.46$]: non-contextual objects (M = 81.92%; SD = 9.21) were recognized more easily than contextual ones (M = 77.19%; SD = 10.96). There also was an effect of emotional value [$F(1,46) = 6.31$, $p<0.05$, $\eta_p^2 = 0.12$]: neutral objects (M = 80.30%; SD = 9.48) were recognized slightly more easily than emotional objects (M = 78.81%; SD = 10.49).

There was an interaction between emotional and contextual values [$F(1,46) = 53.04$, $p<0.001$, $\eta_p^2 = 0.53$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F(1,46) = 49.63$, $p<0.001$, $\eta_p^2 = 0.52$]. Non-contextual neutral objects (M = 85.58%; SD = 7.94) were recognized more easily than non-contextual emotional objects (M = 78.26%; SD = 11.49). Second, there was an effect of emotion on contextual objects as well [$F(1,46) = 20.87$, $p<0.001$, $\eta_p^2 = 0.31$]. Contextual emotional objects (M = 79.36%; SD = 10.31) were recognized more easily than neutral contextual objects (M = 75.02%; SD = 12.45).

Accuracy did not correlate significantly with the presentation of any SF.

Effect of condition on response times

The mean RT for correct positive trials was 623 ms (SD = 83). To analyse possible effects of condition on RTs, without taking SFs into account, we conducted a 2 (Context: non-contextual or contextual) x 2 (Emotion: neutral or emotional) repeated-measures ANOVA on -log(x+1)-transformed RT means per participant (Ratcliff, 1993). Aberrant scores (over 2 seconds) were excluded from the analysis. There was an effect of contextual value on RTs [$F(1,46) = 161.29$, $p<0.001$, $\eta_p^2 = 0.78$] whereby non-contextual objects (Md[1] = 596 ms; SD = 60) were recognized faster than contextual ones (Md = 537 ms; SD = 67). There was no effect of emotional value [$F(1,46) < 1$].

There also was an interaction between emotional value and contextual value [$F(1,46) = 18.46$, $p<0.001$, $\eta_p^2 = 0.29$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F(1,46) = 12.53$, $p<0.001$, $\eta_p^2 = 0.21$]. Non-contextual neutral objects (Md = 532 ms; SD = 57) were identified faster than non-contextual emotional objects (Md = 548 ms; SD = 68). There also was an effect of emotion on contextual objects [$F(1,46) = 10.15$, $p<0.01$, $\eta_p^2 = 0.18$]. Contextual emotional objects (Md = 579 ms; SD = 64) were identified faster than contextual neutral ones (Md = 609 ms; SD = 80).

Effect of spatial frequencies on response time

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between z-scored

[1] Median reaction times are given, since the ANOVA was performed on log transformed

transformed RTs (see Methods: Spatial frequency data analysis) and corresponding sampling vectors for correct positive trials. All object categories confounded, SFs between 13.71 and 24.31 cycles per object width (cpo) correlated negatively with RTs (peak at 19.45 cpo, $Z_{max} = 3.94$, $k = 23$, $p < 0.01$; figure 3a). In other words, RTs were consistently reduced with the presentation of SFs within these boundaries. To examine a possible effect of emotional value, we contrasted classification vectors for all emotional objects and all neutral objects. There was no significant difference ($p > 0.05$). Similarly, there was no significant difference between non-contextual and contextual objects ($p > 0.05$).

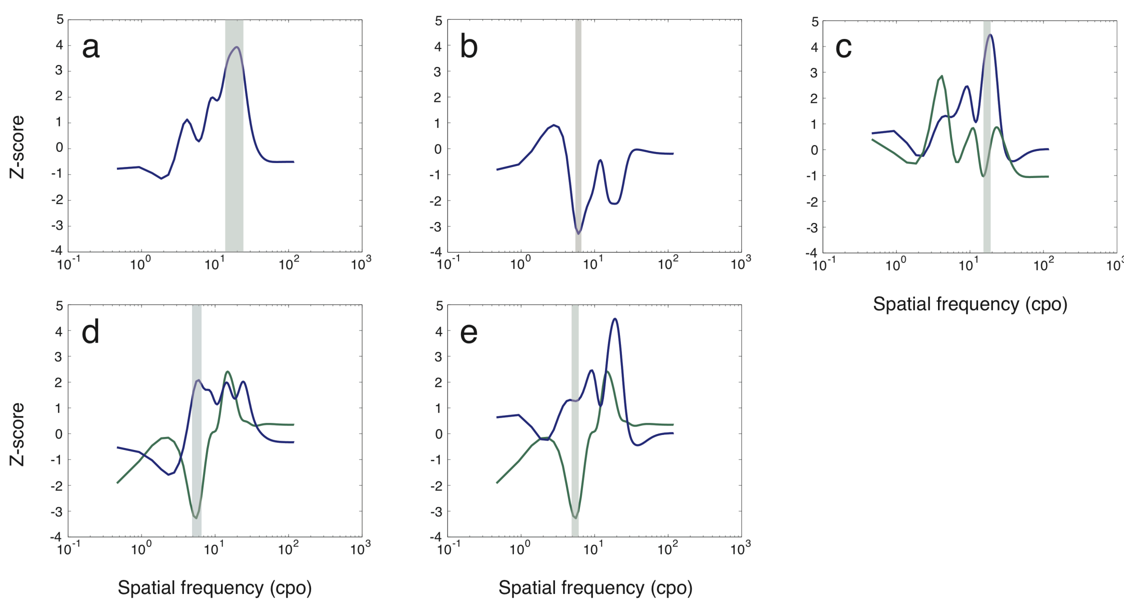

Figure 3. Group classification vectors depicting the correlations between SFs and RTs for different conditions. Higher z-scores indicate a negative correlation (SFs leading to shorter RTs) while lower z-scores indicate a positive correlation (SFs leading to longer RTs). Highlighted grey areas are significant ($p<0.05$). See text for details. **a)** All objects together. **b)** The vector depicting potential interactions between both variables, obtained by contrasting the contrasts of contextual value for both levels of emotional value. **c)** Non-contextual neutral (green) objects and contextual neutral (blue) objects. **d)** Contextual

emotional (green) and non-contextual emotional (blue) objects. **e)** Contextual emotional (green) and contextual neutral (blue) objects.

We then examined the interaction between affective and contextual values (see Methods: Spatial frequency data analysis). We found a significant interaction for SFs between 5.52 and 6.69 cpo (peak at 6.02 cpo, $Z_{max} = 3.29$, $k = 3$, $p < 0.05$; figure 3b).

We subsequently decomposed the interaction into simple effects. There was a significant effect of contextual value on neutral objects between 15.25 and 19.20 cpo; these SFs were correlated more negatively with RTs for contextual neutral objects than for non-contextual neutral objects (peak at 18.98 cpo, $Z_{max} = 3.36$, $k = 9$, $p < 0.05$, corrected for multiple comparisons; figure 3c). However, the interaction was not significant for these SFs, making this effect difficult to interpret. There also was an effect of contextual value for emotional objects: SFs between 4.86 and 6.56 cpo correlated more positively with RTs for contextual emotional objects than for non-contextual emotional objects (peak at 5.56 cpo, $Z_{max} = 3.75$, $k = 4$, $p < 0.05$, corrected for multiple comparisons; figure 3d). Moreover, there was an effect of emotional value on contextual objects: SFs between 4.86 and 6.09 cpo correlated more positively with RTs for contextual emotional objects than for contextual neutral objects (peak at 5.56 cpo, $Z_{max} = 3.21$, $k = 3$, $p < 0.05$, corrected for multiple comparisons; figure 3e). Finally, we observed no significant difference between non-contextual neutral and non-contextual emotional objects ($p > 0.05$). The interaction thus seems to be caused by the significant effect of contextual value on emotional but not on neutral objects, combined with the significant effect of emotional value on contextual but not on non-contextual objects.

Discussion

General spatial frequency use

A few studies have examined the effect of specific SF band filtering during name-picture verification tasks, similar to ours. Collin and McMullen (2005) reported that low-pass filtering objects had little impact on basic-level verification (e.g., verify the “cat” category instead of the “animal” category at the superordinate level or the “Siamese” category at the subordinate level), suggesting that basic-level categorisation does not especially rely on LSFs. Furthermore, Harel and Bentin (2009) reported that basic-level categorisation was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially useful for recognition at the basic level. However, Harel and Bentin's cutoff for HSFs was especially high (65 cpo, or 6.5 cpd), thus preserving only very fine information typically not useful for object recognition. A large band of intermediate spatial frequencies was not explored in these studies.

An important aspect of our study is that instead of applying filters with fixed arbitrary cut-offs, we randomly sampled the entire SF spectrum. This allowed us to overcome the need of selecting arbitrary SF bands to evaluate. Indeed, there is no consensus in the literature about what consists of LSFs or HSFs: this seems to be more understood as a relative measure for SF bands inside a given study. Cut-offs for LSFs in the literature vary from 5 cpo (Boutet, Collin & Faubert, 2003) to 15 cpo (Alorda et al., 2007). Similarly, cut-offs for HSFs vary from 20 cpo (Boutet, Collin & Faubert, 2003) to 65 cpo (Harel & Bentin, 2009). When cut-offs are translated into cycles per degree (cpd), acknowledging that the diagnostic SFs may vary according to viewing distance, the discrepancy is even larger: cut-offs for LSFs vary from less than 0.4 cpd (Boutet, Collin & Faubert, 2003) to more than 2.4 cpd (Alorda et al., 2007) and cut-offs for HSFs vary from 1.4 cpd (Boutet, Collin & Faubert, 2003) to 6.5 cpd (Harel & Bentin, 2009). Quite interestingly, we note that some SFs (between 1.4 and 2.4 cpd) may be included either in LSFs or HSFs.

Our random sampling of the entire SF spectrum allowed us to evaluate the use of SFs considered as neither low nor high by most previous studies. Using this unbiased experimental approach, we found that intermediate SFs between about 14 and 24 cpo (2.3 to 4 cpd) are associated with fast RTs for basic-level verification. This suggests that objects are processed particularly rapidly through these SFs. Although this interpretation is the most straightforward, it is also possible that object processing was at least partly completed before the presentation of the words and, therefore, that the RTs reflect remnants of object processing rather than object processing *per se*.

Another unique aspect of our study is the fact that we equalized SF content of the object images prior to their sampling. This allows us to interpret results more confidently in terms of content of internal representations. Indeed, if SF content is not normalized among stimuli, results most likely reflect an interaction of the stored representation with the information available in the stimulus. Unfortunately, few studies have applied this procedure. As a notable exception, Willenbockel et al. (2010b) did equalize SF spectrum and randomly sample SFs in a face recognition task. Results revealed that SFs peaking at approximately 9 and 13 cycles/face (equivalent to 1.4 and 2 cpd, i.e. SFs that may be categorised as LSFs, HSFs, or most often neither of these) were most correlated with fast and accurate face identification. Although these SFs specific for images of faces are likely to differ from the SF content of object representations, they are an additional indicator that, as in the present study, intermediate SFs rather than LSFs occupy the greatest place in our representation of the world. It is plausible that stored representations consist of mostly these SFs because they are part of the intermediate band of SFs to which we are naturally most sensitive (e.g., Watson & Ahumada, 2005).

Interaction between affective and contextual values

No main effect of contextual or affective value was observed in the SFs correlating with the objects' fast identification. However, we found a significant interaction between affective and contextual values for SFs centered on 6 cpo (or 1 cpd). This indicates that these LSFs, those usually associated with the magnocellular pathway (Derrington & Lennie, 1984), are sensitive in a non-linear manner to a combination of the visual object's intrinsic properties.

When testing the simple effects, we observed that affective value elicited a significant difference in the use of these SFs in contextual objects: they led to longer RTs for contextual emotional objects than for contextual neutral ones. This is not in accordance with the general effect of affective value usually reported in the literature (i.e. LSFs leading to faster RTs; e.g., Mermillod et al., 2010); however, our result is due to an interaction between affective and contextual values and is therefore difficult to compare to those of other studies. Moreover, our stimuli were equalized in their SF content and always comprised several randomly sampled SF bandwidths at the same time, whereas in studies using filters with fixed cut-offs, only some specific band of LSFs or HSFs is shown at a time.

SFs near 6 cpo (or 1 cpd) also led to longer RTs for contextual emotional objects than for non-contextual emotional objects. The effect of contextual value on SF content of object representations had not been tested before but it had been often proposed that rapidly extracted

LSFs are sufficient to activate representations associated with an object (Bar, 2004; Fenske et al., 2006). Our data suggest that these presumed/hypothetical associative representations do not speed up the object's recognition. Why we observed this modulation only for emotional objects is not clear, but several interactions between affective and contextual processing have already been reported and could possibly explain the discrepancy (Brunyé et al., 2013; Shenhav et al., 2013; Storbeck & Clore, 2005). For example, affective value might influence the extent to which we associate a particular object to other objects (Bar, 2009; Shenhav et al., 2013).

Conclusion

The main findings of the present study are i) that the SF content of object representations in general are in an intermediate band between 14 and 24 cpo (2.3 to 4 cpd), and ii) that intrinsic high-level categorical properties of an object influence the SF content of its internally stored representation, more precisely that affective and contextual values interact in their modulation of the SF content of object representations.

According to predictive accounts of brain function (e.g., Bar, 2003; Rao & Ballard, 1999; Friston, 2003, 2010; Friston et al., 2006), our mind constantly generates predictions about our environment, and our understanding of a sensory input is based both on the available sensory information and on prior beliefs stored as internal representations (see Knill & Pouget, 2004). In this study, we investigated precisely the SF content of these stored representations, and its potential flexible modulation by affective and contextual properties of the stimulus. Our results reveal that stored representations of visual objects are composed of intermediate SFs that are often left over in studies using filters with fixed arbitrary cut-offs. Furthermore, we observed a modulation of this SF content by affective and contextual intrinsic values of the visual object, suggesting its flexibility and thus the multiplicity of visual recognition systems.

Our study cannot however address directly the issue of temporal dynamics of visual object recognition. While we observed that some SFs are more useful to identify some objects, we cannot conclude that these are extracted first. Further studies should address these issues and their links to potential initiation of top-down mechanisms.

References

Alorda, C., Serrano-Pedraza, I., Campos-Bueno, J. J., Sierra-Vázquez, V., & Montoya, P. (2007). Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia*, *45*(14), 3223–3233. doi:10.1016/ j.neuropsychologia.2007.06.017

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*(7), 1493–1503. doi:10.1093/cercor/bhl078

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. doi:10.1126/ science.8316836

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. doi:10.1038/nrn1476

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 1181–1182.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449-454.

Barceló, F., Suwazono, S., & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nature Neuroscience*, *3*(4), 399–403. doi:10.1038/73975

Barrett, L. F., & Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1325–1334. doi:10.1093/brain/106.2.473

Biederman, I. (1972). Perceiving Real-World Scenes. *Science*, *177*(4043), 77–80.

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception*, *10*(5), 573–578.

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

Bocanegra, B. R., & Zeelenberg, R. (2009). Emotion improves and impairs early vision. *Psychological Science*, *20*(6), 707–713. doi:10.1111/j.1467-9280.2009.02354.x

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception & Psychophysics, 65*(7), 1078-1093.

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology*, *46*(1), 1–11. doi:10.1111/j.1469-8986.2008.00702.x

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *Journal of Neuroscience*, *22*(5), 1976–1984.

Brunyé, T. T., Gagnon, S. A., Paczynski, M., Shenhav, A., Mahoney, C. R., & Taylor, H. A. (2013). Happiness by association: breadth of free association influences affective states. *Cognition*, *127*(1), 93–98. doi:10.1016/j.cognition.2012.11.015

Bullier, J. (2001). Integrated model of visual processing, *Brain Research Reviews, 36,* 96–107.

Calderone, D. J., Hoptman, M. J., Martínez, A., Nair-Collins, S., Mauro, C. J., Bar, M., et al. (2013). Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cerebral Cortex*, *23*(8), 1849–1858. doi:10.1093/cercor/bhs169

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 659–667.

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorisation relies on high spatial frequencies to a greater degree than basic-level categorisation, *Perception & Psychophysics, 67*(2), 354–364.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial Vision*. New York: Oxford University Press.

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*, 219–240.

Efron, B., & Tibshirani, R. (1993). A*n Introduction to the Bootstrap.* London: Chapman & Hall/CRC.

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research*, *155*, 3–21. doi:10.1016/S0079-6123(06)55001-0

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. doi:10.1016/j.neunet.2003.06.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 607–608. doi:10.1038/nrn2787-c2

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology - Paris*, *100*(1-3), 70–87. doi:10.1016/j.jphysparis.2006.10.001

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677–696. doi:10.1016/j.neuron.2007.05.019

Grossberg, S. (1980). Biological competition: Decision rules, pattern formation, and oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(4), 2338–2342.

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1264–1273. doi:10.1037/a0013621

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439. doi:10.1016/j.pneurobio.2007.09.001

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*(4), 800-802.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. doi:10.1016/j.tins.2004.10.007

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232–13240. doi:10.1523/JNEUROSCI.3481-07.2007

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Frontiers in Psychology, 3:*107. doi:10.3389/fpsyg.2012.00107

Logothetis, N. K., & Sheinberg D. L. (1996) Visual object recognition. *Annual Review of Neuroscience*, *19,* 577-621.

Maunsell, J., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*, 363-401.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain & Cognition*, *57*(2), 151–157. doi:10.1016/j.bandc.2004.08.035

Mermillod, M., Droit-Volet, S., Devaux, D., Schaefer, A., & Vermeulen, N. (2010). Are coarse scales sufficient for fast detection of visual threat? *Psychological Science*, *21*(10), 1429–1437. doi:10.1177/0956797610381503

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS One*, *7*(6), e38493. doi:10.1371/journal.pone.0038493

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512. doi:10.1126/science.1057099

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Ratcliff, R. (1993). Methods for dealing with reaction outliers. *Psychological Bulletin, 114*(3), 510-532.

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dpendent Scene Recognition. *Psychological Science*, *5*(4), 195–200.

Shenhav, A., Barrett, L., & Bar, M. (2013). Affective value and associative processing share a cortical substrate. *Cognitive, Affective, and Behavioural Neuroscience, 13,* 46-59.

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: mood and the false memory effect. *Psychological Science*, *16*(10), 785–791. doi:10.1111/j.1467-9280.2005.01615.x

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 1–11.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6*(6), 624-631.

Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9), 6–6. doi:10.1167/5.9.6

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A, 4*(10), 2006-2021.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010a). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. doi:10.3758/BRM.42.3.671

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010b). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 122–135. doi:10.1037/a0016465

Willenbockel, V., Lepore, F., Nguyen, D.K., Bouthillier, A., & Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions — an intracranial ERP study. *Frontiers in Psychology, 3:*237, doi: 10.3389/fpsyg.2012.00237

Supplementary Material

List of the stimuli, along with the object category in which they were classified following our validation. The object names are the ones that were presented on the screen. We also provide an English translation.

Object Category	**Object Name**	**English Translation**
Contextual Emotional	Berceau	Cradle
	Bonhomme de neige	Snowman
	Boule disco	Disco Ball
	Brancard	Stretcher
	Cadeau	Gift
	Casque de soldat	Soldier's Helmet
	Chaise longue	Deck Chair
	Chaise électrique	Electric Chair
	Fauteuil de dentiste	Dentist's Chair
	Gâteau d'anniversaire	Birthday Cake
	Hamac	Hammock
	Masque et tuba	Snorkeling Mask
	Mitraillette	Machine Gun
	Palmes	Diving Fins
	Pistolet	Handgun
	Seau de plage	Beach Pail
	Table de billard	Pool Table
	Tombe	Tombstone
Non-contextual Emotional	Poubelle	Trash Can
	Verre à cocktail	Cocktail Glass
	Araignée	Spider
	Billet de banque	Dollar Bill
	Bombe	Bomb
	Bonbon	Candy
	Bouquet de fleurs	Flower Bouquet
	Cadre pour photo	Picture Frame
	Carré d'as	Four of a Kind

	Crâne	Skull
	Dynamite	Dynamite
	Fauteuil roulant	Wheelchair
	Grenade	Grenade
	Téléphone cassé	Broken Phone
	Papillon	Butterfly
	Scie mécanique	Chainsaw
	Tirelire	Piggy Bank
	Verre de bière	Glass of Beer
Contextual Neutral	Baignoire	Bathtub
	Casque de sécurité	Safety Helmet
	Chariot d'épicerie	Grocery Cart
	Cuisinière	Oven
	Feu de circulation	Traffic Light
	Gouvernail	Rudder
	Marteau de tribunal	Gavel
	Pompe à essence	Fuel Pump
	Roulette de casino	Roulette
	Réfrigérateur	Refrigerator
	Satellite	Satellite
	Sièges d'avion	Aircraft Seats
	Sombrero	Sombrero
	Tampon et encrier	Stamp and Ink
	Tente	Tent
	Tracteur	Tractor
	Ventilateur	Fan
	Voiturette de golf	Golf Cart
Non-contextual Neutral	Appareil photo	Camera
	Banc	Bench
	Bouteille	Bottle
	Caisse en plastique	Plastic Crate
	Chaussure	Shoe
	Coussin	Cushion
	Horloge	Clock
	Jumelles	Binoculars
	Lampe de poche	Flashlight
	Livre	Book
	Mouchoirs	Tissues
	Panier	Basket
	Pelote de ficelle	Ball of String

Prise électrique	Electrical Outlet
T-shirt	T-Shirt
Téléphone portable	Mobile Phone
Téléphone	Phone
Étagère	Shelf