



HAL
open science

Reconnaissance visuelle : Etude de l'influence de la vision périphérique dans le cadre d'un modèle fréquentiel prédictif

Alexia Roux-Sibilon

► **To cite this version:**

Alexia Roux-Sibilon. Reconnaissance visuelle : Etude de l'influence de la vision périphérique dans le cadre d'un modèle fréquentiel prédictif. Psychologie. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALS029 . tel-03247300

HAL Id: tel-03247300

<https://theses.hal.science/tel-03247300>

Submitted on 3 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **PCN – Sciences cognitives, psychologie et neurocognition**

Arrêté ministériel : 25 mai 2016

Présentée par

Alexia ROUX-SIBILON

Thèse dirigée par **Carole Peyrin**, Université Grenoble Alpes

préparée au sein du **Laboratoire de Psychologie et Neurocognition**

dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Reconnaissance visuelle : Etude de l'influence de la vision périphérique dans le cadre d'un modèle fréquentiel prédictif

Thèse soutenue publiquement le **10 juillet 2020**,
devant le jury composé de :

Madame CAROLE PEYRIN

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES,
Directrice de thèse

Monsieur NICOLAS POIREL

PROFESSEUR DES UNIVERSITES, UNIVERSITE PARIS 5, Rapporteur

Monsieur QUENTIN LENOBLE

MAITRE DE CONFERENCES HDR, UNIVERSITE DE LILLE, Rapporteur

Madame VALERIE GOFFAUX

DOCTEUR EN SCIENCES, UNIVERSITE CATHOLIQUE LOUVAIN –
Belgique, Examinatrice

Monsieur MARTIAL MERMILLOD

PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES,
Président



Résumé

Nous saisissons l'essentiel de notre environnement visuel en un seul coup d'œil, et ce indépendamment de sa complexité. Différentes propositions théoriques dans le domaine de la reconnaissance visuelle ont proposé que cette capacité reposerait sur l'extraction du *gist* de la scène visuelle, une première représentation visuelle globale et rudimentaire. L'extraction du *gist* reposerait notamment sur l'analyse rapide des basses fréquences spatiales du signal visuel, qui permettrait une catégorisation assez grossière de la scène et des objets qui la composent. Cette première représentation de la scène permettrait par ailleurs de déclencher des mécanismes prédictifs qui guideraient dans un second temps une analyse visuelle plus minutieuse et détaillée. Ce cadre théorique, que nous avons formalisé dans cette thèse sous le nom de modèle fréquentiel prédictif, rend compte d'un grand nombre de résultats empiriques. Dans l'ensemble, les travaux à l'appui de ce modèle se sont concentrés sur la reconnaissance visuelle de petits stimuli présentés en vision centrale. Or, l'acuité visuelle diminue drastiquement avec l'excentricité rétinienne si bien qu'une grosse part du signal en basses fréquences spatiales provient de la vision périphérique. Dans ce travail de thèse, nous avons testé l'hypothèse que l'information de basse résolution spatiale extraite en vision périphérique influencerait les traitements réalisés en vision centrale. Dans une première série d'expériences, nous avons observé que la catégorisation de stimuli visuels complexes (de type objets ou scènes) présentés en vision centrale peut être améliorée et/ou accélérée lorsque l'information visuelle périphérique est sémantiquement congruente par rapport à lorsqu'elle est incongruente. Dans une seconde série d'expériences, nous avons étudié les conséquences d'une perte de vision périphérique sur les traitements réalisés en vision centrale. Nous avons pour cela testé des patients atteints d'un glaucome. Cette pathologie ophtalmique affecte en particulier la rétine périphérique et représente ainsi un bon modèle pathologique d'un système dans lequel la vision périphérique a moins d'importance. Nous avons observé que ces patients présentaient un déficit pour catégoriser des petites scènes présentées en vision centrale, alors même que leur sensibilité rétinienne était préservée pour cette région du champ visuel. Ce dernier résultat suggère que la perte de vision périphérique

causée par le glaucome désorganise les mécanismes prédictifs et perturberait ainsi la reconnaissance visuelle de manière globale, même en vision centrale. L'ensemble de ces travaux de thèse nous permettent de conforter et préciser le modèle fréquentiel prédictif tout en soulignant l'importance de considérer la vision périphérique dans les modèles de reconnaissance visuelle.

Abstract

We understand most of our visual environment at a glance, regardless of its complexity. Theories of visual recognition postulate that this ability is based on the extraction of the gist of the visual scene, a first global and coarse visual representation. Gist perception would be based on the rapid analysis of low spatial frequencies in the visual signal, and would allow a coarse categorization of the scene and the objects that compose it. This first representation of the scene would also trigger predictive mechanisms that guide a more detailed visual analysis. This theoretical framework, which we formalized in this thesis under the name of predictive frequency model, accounts for a large number of empirical results. Overall, work in support of this model has focused on the visual recognition of small stimuli presented in central vision. However, visual acuity decreases drastically with retinal eccentricity so that a large part of the signal at low spatial frequencies comes from peripheral vision. In this doctoral thesis, we tested the hypothesis that the low spatial resolution information extracted in peripheral vision would influence the processing of detailed information in central vision. In a first series of experiments, we observed that the categorization of complex visual stimuli (objects or scenes) presented in central vision can be improved and / or accelerated when information in peripheral vision is semantically congruent compared to when it is incongruent. In a second series of experiments, we studied the consequences of a peripheral vision loss onto visual recognition in central vision. We tested patients with glaucoma. This ophthalmological condition affects the peripheral retina, thus representing a good pathological model of a system in which peripheral vision is degraded. We observed that these patients presented a deficit to categorize small scenes presented in central vision, even though their retinal sensitivity was preserved for this region of the visual field. This last result suggests that the loss of peripheral vision disrupts the predictive visual system and would thus disturb visual recognition globally, even in central vision. Overall, the results allow consolidating and clarifying the predictive frequency model while emphasizing the relevance of considering peripheral vision in visual recognition models.

Remerciements

Je remercie très chaleureusement Nicolas Poirel, Quentin Lenoble, Valerie Goffaux et Martial Mermillod d’avoir accepté d’évaluer cette thèse. Je remercie aussi Nicolas Poirel, ainsi que Sylviane Valdois, pour avoir suivi le déroulement de mes travaux de thèse par le biais du comité de suivi individuel.

Cette thèse est l’aboutissement d’un long travail qui a impliqué beaucoup plus d’une personne. Je veux remercier ici toutes les personnes qui d’une façon ou d’une autre se sont retrouvées entre moi et mon doctorat.

Je tiens d’abord à exprimer toute ma reconnaissance à ma directrice de thèse Carole Peyrin. Cela a vraiment été une chance d’avoir pu bénéficier de ta formation et de ta pédagogie. Merci pour ton soutien, pour ta clairvoyance et ta perspicacité qui m’ont débloquée nombre de fois, mais aussi, tout particulièrement, pour le *temps* que tu m’as accordé, sans exception, sans condition, et ce depuis les réunions de travail, lorsque j’étais en master 1, où tu m’expliquais comment faire une fonction IF sur Excel jusqu’à la toute fin de rédaction de cette thèse – il n’y a pas un moment où tu n’es pas disponible. Je suis fier d’avoir pu contribuer à tes recherches, merci de ta confiance.

Merci à l’Université Grenoble Alpes (bourse AGIR POLE CBS) et la Fondation de France (bourse Berthe Fouassier) qui ont financé ces recherches.

Je veux remercier également l’équipe de la clinique d’ophtalmologie du CHUGA, notamment le Pr. Florent Aptel et le Pr. Christophe Chiquet pour leur accueil dans le service et leur collaboration dans les projets qui ont impliqué les patient-es atteints d’un glaucome. Merci tout particulier à Floriane Rutgé, qui a réalisé son projet de recherche de M2 neurosciences cliniques à nos côtés durant son internat de médecine, et qui a été d’une aide considérable pour le recrutement des patient-es. Merci pour ta pédagogie et les échanges enrichissants qu’on a pu avoir ensemble. Ce travail aurait été beaucoup plus difficile sans toi !

Ma gratitude va également à tous les étudiants et étudiantes qui ont participé aux expériences de ce travail de thèse ainsi qu’aux patient-es de la clinique d’ophtalmologie du CHUGA.

Merci aux ancien-nes qui m’ont précédée dans l’équipe de Carole, notamment à Stephen Ramanoël et à Louise Kauffmann. Merci surtout Louise pour ton aide en continu pendant toute ma thèse, pour l’énergie que tu avais mise dans l’article Visual Cognition à l’époque, pour ton aide sur les détails méthodo, et pour les discussions sur la théorie (sharpening dampening). Mais surtout, soyons honnêtes, pour les pauses cigarette.

Merci à Nathalie pour tes encouragements et ta bienveillance, et pour ton aide pour le traitement d’image. Je remercie aussi la plateforme IRMage et la cellule de neuroimagerie fonctionnelle du laboratoire, Emilie, Laurent, et surtout Cédric qui a

passé du temps à me former au traitement des données IRMf. Vu mon enthousiasme et mes capacités à traiter ce genre de données, tu as été d'une patience incroyable. Merci pour tous tes conseils précieux.

Merci aux étudiants qui ont contribué à l'acquisition des données, spécialement à Dorante, Malena et Sarah. Et Audrey bien sûr, je te remercie vraiment pour ton soutien et pour avoir partagé avec moi le désarroi de devoir mettre de côté les données de neuf patients sur dix à cause des critères d'inclusion. C'était pas facile !

Je tiens aussi à remercier chaleureusement le LPNC et Monica Baciou pour avoir permis que cette thèse se réalise dans de si bonnes conditions. Je veux exprimer ma reconnaissance aux membres du laboratoire. À mes ancien·nes profs de licence ou de master, Elsa, Aurélie, Alan, Martial, Eve, Karine, Stéphane, Richard, David, Hélène, Nathalie, ... Vous m'avez fait adorer la psychologie expérimentale et ce n'est pas pour rien que j'ai voulu faire de la recherche, c'est une chance d'avoir pu suivre vos cours. Mais merci aussi à tous·tes les autres. Puisque l'enseignement m'a offert des bouffées d'air frais pendant ces années de recherche, je remercie également l'UFR et le Département de Psychologie, et les responsables d'UE qui m'ont accordé leur confiance, même aux débuts, Marcela, Martial, David, Aurélie, Richard, Jean-Charles.

À mes copines de thèse Emilie, Audrey, Lisa et Hélène. C'était une drôle d'aventure qu'on a toutes menée différemment, j'ai hâte de voir ce que le futur nous réserve à toutes les cinq. Quand c'était difficile vous avez rendu les choses drôles, et j'ai énormément appris de vous. Alors merci pour tout les filles.

Merci aux autres copines et copains du (des) labo(s), notamment à Cédric, compagnon d'ATER-galère, à Ladislav (bien heureuse maintenant que tu m'aies convaincue de faire des modèles mixtes pour ces données de thèse!) – je pense à vous deux et Audrey et ce long mois d'août au Nostrum à écrire ces thèses... plus qu'une et on est bon!, à Amélie et à Brice, qui sont trop loin! à Rémi co-capitaine, à Morgane, Kevin, Svetlana, Chloé, Adeline, Léa, Elie, Brice, Céline, Elise, Merrick, Sonja, Elena, Laura, et aux autres que j'ai l'air d'oublier mais que je n'oublie pas! merci à l'énergie et à l'humour des nouveaux arrivants qui rendent plus supportables les fins de thèse. Bref, merci à vous tous·tes pour les coups de main, le soutien, les bières au Martin's le vendredi soir et le reste. Et merci les Furies pour les entraînements et les matchs qui nous sortent un peu le nez du labo !

Enfin, je remercie avec émotion mes amis de toujours Sandra, Thomas, Benoît, Léa, Marion, Mikael, Xavier, Sébastien, Alban pour leur amitié et leur soutien. Pardon d'avoir été de moins en moins disponible ces dernières années, c'est à cause du document que vous avez sous les yeux. Merci Lélé et Hug pour les corrections de dernière minute, vous êtes désormais responsables de tout problème d'orthographe qu'il pourrait y avoir dans cette thèse. Merci Sand et Jo d'avoir amené peu de légèreté pendant cette période de ma vie grâce à l'arrivée d'Artus puis de Marjane. Merci aux copines et copains de Grenoble et d'ailleurs. Et merci à ma famille, je sais que ce n'est pas toujours très clair pour vous, mais malgré tout vous m'avez toujours soutenue dans mes choix. Et les

portes sont toujours ouvertes quand j'ai besoin de réconfort ! J'ai alors une grosse pensée pour mon papa, ma maman, ma sœur Laura et mon frère Alan, Jean et Yvonne mes grands-parents exceptionnels, Slavica ma mamie que j'adore – et mon papy Miki (qui m'aurait dit 'maintenant, bientôt mariée?'), mes tantes et mes oncles, mes cousines et mes cousins, mais aussi à ma belle-famille Sylvie, Angelo, Romy. Vous allez toutes et tous beaucoup me manquer quand je serai en Belgique.

Et par-dessus tout merci mon Hugo – des flous aux basses fréquences, des hautes fréquences aux normal maps, on trouvera bien toujours le moyen de s'entendre !

Table des matières

<i>Résumé</i>	v
<i>Abstract</i>	vi
<i>Remerciements</i>	vii
<i>Chapitre 1 Cadre théorique : Modélisation de la reconnaissance visuelle</i>	3
1 <i>Reconnaissance visuelle ultra-rapide</i>	4
1.1 Définitions : reconnaissance et catégorisation visuelle	4
1.2 Évaluation de la reconnaissance visuelle ultra-rapide	5
1.3 La perception de l'essentiel : le concept de <i>gist</i> de la scène	8
1.3.1 Contenu perceptuel du <i>gist</i> : rôle des basses fréquences spatiales et des <i>summary statistics</i>	8
1.3.2 Contenu conceptuel du <i>gist</i> : Quel niveau de catégorisation au premier coup d'œil?	12
2 <i>Du modèle fréquentiel à un modèle fréquentiel prédictif de la reconnaissance visuelle</i>	16
2.1 Le système visuel : un analyseur de Fourier	16
2.1.1 Les fréquences spatiales	16
2.1.2 Un outil pour étudier le rôle des fréquences spatiales dans la perception visuelle : Le filtrage spatial	17
2.1.3 Bases biologiques de l'analyse fréquentielle de l'information visuelle : Les systèmes magnocellulaire et parvocellulaire	18
2.2 Le modèle fréquentiel	21
2.2.1 Le phénomène de précedence globale et les formes hiérarchiques	21
2.2.2 Précedence du traitement des BFs dans la perception de réseaux sinusoïdaux et de scènes	24
2.2.3 Limites du modèle fréquentiel : Un principe par défaut	26
2.3 Influences descendantes et mécanismes prédictifs du traitement visuel	27
2.3.1 Le cortex visuel : Une architecture hiérarchique mais des connexions bidirectionnelles	27

2.3.2	Codage prédictif et modèles neurocognitifs prédictifs de la reconnaissance visuelle	30
2.4	Modèle fréquentiel prédictif	36
2.5	Importance d'un test direct de l'influence des BF's sur le traitement des HF's	37
3	<i>Limite du modèle Fréquentiel Prédicatif : Qu'en est-il de la vision périphérique ?</i>	43
3.1	Une correspondance directe entre le concept de <i>gist</i> et le contenu visuel périphérique	43
3.1.1	Résolution spatiale non homogène du champ visuel	43
3.1.2	Contribution de la vision périphérique dans la perception du <i>gist</i> d'une scène	45
3.1.3	<i>Crowding</i> et <i>summary statistics</i> en vision périphérique	47
3.2	Arguments en faveur de l'hypothèse d'une influence de la vision périphérique sur la vision centrale	49
4	<i>Objectif des travaux de thèse et méthodologie générale</i>	54
4.1	Paramètres expérimentaux généralement adoptés	54
4.2	Objectif : Test d'un modèle FP-peri	54
4.3	Tâches de catégorisation	57
4.4	Approche méthodologique chez le sujet jeune sain : Le paradigme d'interférence sémantique	58
4.5	Le glaucome comme modèle pathologique d'une perturbation de la vision périphérique	59
 <i>Chapitre 2 Contribution empirique :</i>		
<i>Influence de la vision périphérique sur les traitements réalisés en vision centrale chez le sujet sain</i> 63		
1	<i>Préambule pour le traitement des données et l'analyse statistique</i>	64
1.1	Analyse de la proportion de réponses correctes	65
1.2	Analyse des temps de réponses correctes	65
2	<i>Expérience 1 - Étude de l'intégration des basses et hautes fréquences spatiales dans des scènes hybrides</i>	67
2.1	Objectifs et hypothèses	67
2.1.1	Précédence temporelle du traitement des BF's lors de la catégorisation de scènes, et rôle du temps de traitement	68
2.1.2	Influences mutuelles des BF's et des HF's lors de la catégorisation des scènes hybrides	69
2.1.3	Nature des influences : Effet de la similarité physique	69
2.2	Méthode	70
2.2.1	Participants	70
2.2.2	Stimuli et procédure	71
2.3	Résultats	74

2.3.1	Session 1 : Catégorisation des scènes filtrées	74
2.3.2	Session 2 : Catégorisation des scènes hybrides	76
2.4	Discussion	79
3	<i>Expériences 2 et 3. Influence d'une scène en vision périphérique sur la catégorisation de scènes en vision centrale</i>	86
3.1	Objectifs et hypothèses	86
3.1.1	Test d'une influence de l'information en BF _s extraite en vision périphérique sur la catégorisation de scènes en vision centrale	87
3.1.2	Nature de l'influence prédictive périphérique : Influence de la similarité physique	87
3.1.3	Bases cérébrales de l'interférence périphérique	88
3.2	Expérience 2 : Méthode	89
3.2.1	Participants	89
3.2.2	Stimuli et procédure	89
3.3	Expérience 2 : Résultats	91
3.3.1	Analyse de la proportion de réponses correctes	91
3.3.2	Analyse des temps de réponses correctes	92
3.4	Expérience 3 : Méthode	93
3.4.1	Participants	93
3.4.2	Stimuli et procédure	93
3.4.3	Acquisition des données IRMf	94
3.4.4	Traitement et analyse statistique des données d'IRMf	94
3.5	Expérience 3 : Résultats	95
3.5.1	Analyse de la proportion de réponses correctes	95
3.5.2	Analyse des temps de réponses correctes	96
3.5.3	Analyse des données IRMf	96
3.6	Discussion des Expériences 2 et 3	97
4	<i>Expériences 4 et 5 : Influence de l'information contextuelle d'une scène en vision périphérique sur la catégorisation d'un objet en vision centrale</i>	103
4.1	Objectifs et hypothèses de l'Expérience 4	103
4.2	Expérience 4 : Méthode	105
4.2.1	Participants	105
4.2.2	Stimuli	105
4.2.3	Procédure	106
4.3	Expérience 4 : Résultats	107
4.3.1	Analyse de la proportion de réponses correctes	108
4.3.2	Analyse des temps de réponses correctes	108
4.4	Résumé et limites de l'Expérience 4	108
4.5	Expérience 5 : Objectifs et hypothèses	110
4.5.1	Mesure du seuil de visibilité	111

4.5.2	Direction de l'effet de congruence	111
4.5.3	Contrôle expérimental : estimation de l'importance des réponses précipitées	112
4.5.4	Nature de l'influence prédictive	112
4.6	Expérience 5 : Méthode	113
4.6.1	Participants	113
4.6.2	Stimuli	114
4.6.3	Procédure	116
4.6.4	Traitement des données	117
4.7	Expérience 5 : Résultats	118
4.7.1	Effet de congruence et nature des influences : scènes in- tactes <i>vs</i> scrambled	118
4.7.2	Direction de l'effet de congruence : Comparaison avec la ligne de base	120
4.7.3	Relation entre l'effet de congruence et l'effet des réponses 'précipitées'	120
4.8	Discussion des Expériences 4 et 5	121

*Chapitre 3 Contribution empirique :
Effet d'une dégradation de la vision périphérique sur
les traitements réalisés en vision centrale chez le pa-
tient atteint de glaucome* **127**

1	<i>Contexte et hypothèses</i>	128
1.1	Atteinte de la vision périphérique dans le glaucome	128
1.2	Troubles visuels cognitifs en vision résiduelle	130
1.3	Objectifs et hypothèses théoriques	133
2	<i>Expérience 6 : Étude de la reconnaissance de scènes et de visages faiblement contrastés dans la vision centrale de patients atteints d'un glaucome</i>	136
2.1	Objectifs et hypothèses opérationnelles	136
2.1.1	Participants	138
2.1.2	Stimuli et procédure	139
2.1.3	Analyse des données	143
2.2	Résultats	144
2.2.1	Session expérimentale 'SCENE'	144
2.2.2	Session expérimentale 'VISAGE'	146
2.3	Discussion	148
3	<i>Expérience 7 : Étude exploratoire en IRMf d'une désorganisation fonctionnelle cérébrale consécutive à une atteinte rétinienne glauc- mateuse</i>	152
3.1	Objectifs et hypothèses opérationnelles	152
3.2	Méthode	153

3.2.1	Participants	153
3.2.2	Stimuli et procédure	154
3.2.3	Acquisition et traitement des données IRMf	156
3.3	Résultats	157
3.3.1	Performances comportementales	157
3.3.2	Analyse des données IRMf	158
3.4	Discussion	159
<i>Chapitre 4 Discussion générale</i>		165
1	<i>Apports des résultats à la théorie</i>	167
1.1	Synthèse des résultats obtenus	167
1.2	Taille de l'effet d'interférence périphérique et importance pratique	169
1.3	Limites méthodologiques	172
1.4	Apports des résultats pour les théories prédictives de la reconnaissance visuelle	174
1.4.1	Arguments empiriques d'une influence prédictive du contenu périphérique	174
1.4.2	Nature et mécanismes des prédictions	175
1.4.3	Le cortex orbitofrontal comme site cortical d'un système de prédiction	180
1.4.4	Un système prédictif moins efficace lorsque la vision périphérique est dégradée	183
2	<i>Perspectives théoriques et implications cliniques</i>	185
2.1	Un modèle basé sur la perception du <i>gist</i> doit-il être absolument un modèle fréquentiel?	185
2.2	Modes de vision, modèle de hiérarchie inverse et implications pour les personnes atteintes d'un glaucome	189
<i>Chapitre A Annexe 1</i>		195
<i>Chapitre B Annexe 2</i>		211
<i>Chapitre C Annexe 3</i>		229
<i>Chapitre D Annexe 4</i>		249

Table des figures

Figure 1.1	Image naturelle constituée d’objets non identifiables	5
Figure 1.2	Exemples d’images hybrides	10
Figure 1.3	Représentation schématique de la procédure expérimentale et des stimuli utilisés par Ariely (2001)	11
Figure 1.4	Procédure expérimentale et stimuli de l’étude de de Fockert et Wolfenstein (2009)	12
Figure 1.5	Illustration de l’information contenue dans les basses et hautes fréquences spatiales d’une scène naturelle	18
Figure 1.6	Illustration de la procédure de filtrage spatial d’une image	19
Figure 1.7	Stimuli hiérarchiques	23
Figure 1.8	Réseaux sinusoïdaux	24
Figure 1.9	Latences de réponse visuelle de neurones dans différentes aires corticales (Bullier, 2001)	29
Figure 1.10	Triangle de Kanizsa	31
Figure 1.11	Modèle proactif (Bar, 2003 ; Bar et al., 2006)	32
Figure 1.12	Résultats des analyses de synchronie de phase de Bar et al. (2006)	33
Figure 1.13	Modèle Fréquentiel Prédicatif	37
Figure 1.14	Stimuli et résultats de l’étude de Mu et Li (2013)	39
Figure 1.15	Représentation schématique de la résolution spatiale non uniforme du champ visuel	44
Figure 1.16	Ratio P/M dans la rétine et dans les corps genouillés latéraux (tiré de Azzopardi et al., 1999)	45
Figure 1.17	Paradigme expérimental <i>window-scotoma</i> (Larson & Loschky, 2009)	47
Figure 1.18	Illustration du phénomène de <i>crowding</i> en vision périphérique	48
Figure 1.19	Effets de contexte	51
Figure 1.20	Modèle Fréquentiel Prédicatif intégrant la vision périphérique (modèle FP-peri)	56
Figure 1.21	Modèle FP-peri perturbé par l’atteinte rétinienne glaucomateuse	61
Figure 2.1	Représentation schématique du <i>odds ratio</i>	66
Figure 2.2	Types d’images hybrides utilisées dans l’Expérience 1.	68
Figure 2.3	Exemples de stimuli utilisés dans l’Expérience 1	73
Figure 2.4	Schéma de la procédure expérimentale de l’Expérience 1.	74
Figure 2.5	Résultats de l’Expérience 1 (Session 1)	75
Figure 2.6	Comportement des groupes cible-BFs et cible-HFs (Session 1)	76

Figure 2.7	Résultats de l'Expérience 1 (Session 2) - proportion de réponses correctes	78
Figure 2.8	Résultats de l'Expérience 1 (Session 2) - TRm	80
Figure 2.9	Exemples de stimuli utilisés dans l'Expérience 2.	87
Figure 2.10	Types de stimuli utilisés dans les Expériences 2 et 3	90
Figure 2.11	Schéma de la procédure expérimentale de l'Expérience 2.	91
Figure 2.12	Résultats de l'Expérience 2	92
Figure 2.13	Résultats de l'Expérience 3	97
Figure 2.14	Résultats d'IRMf de l'Expérience 3	98
Figure 2.15	Stimuli de l'Expérience 4	104
Figure 2.16	Procédure expérimentale de l'Expérience 4	107
Figure 2.17	Résultats de l'Expérience 4	109
Figure 2.18	Modulation paramétrique de la cohérence de phase des objets dans l'Expérience 5	111
Figure 2.19	Stimuli utilisés dans l'Expérience 5 (1)	114
Figure 2.20	Stimuli utilisés dans l'Expérience 5 (2)	115
Figure 2.21	Procédure expérimentale de l'Expérience 5	116
Figure 2.22	Résultats de l'Expérience 5	119
Figure 2.23	Résultats de l'Expérience 5. Comparaison avec la ligne de base.	121
Figure 2.24	Résultats de l'Expérience 5 Réponses précipitées	122
Figure 2.25	Transformation difféomorphique de scènes (Stojanoski & Cusack, 2014)	126
Figure 3.1	Réalisation d'une périmétrie statique automatisée.	129
Figure 3.2	Exemple de champs visuels monoculaires (OG : oeil gauche; OD : oeil droit) de patients atteints de glaucome	131
Figure 3.3	Illustration du paradigme des <i>Steady-Pulsed pedestal</i>	132
Figure 3.4	Stimuli utilisés pour l'étude des processus visuels cognitifs de patients atteints de glaucome	134
Figure 3.5	<i>Probability plots</i> des déviations individuelles monoculaires des patients GPAO inclus dans l'Expérience 6.	140
Figure 3.6	Exemples de stimuli utilisés dans l'Expérience 6.	141
Figure 3.7	Résultats de l'Expérience 6 (session SCENE)	145
Figure 3.8	Résultats de l'Expérience 6 (session VISAGE)	147
Figure 3.9	Champs visuels des patients testés dans l'Expérience 7	154
Figure 3.10	Stimuli utilisés dans l'Expérience 7	155
Figure 3.11	Résultats comportementaux de l'Expérience 7	158
Figure 3.12	Résultats d'IRMf de l'Expérience 7 (1)	160
Figure 3.13	Résultats d'IRMf de l'Expérience 7 (2)	161
Figure 4.1	Proposition d'un protocole expérimental	171
Figure 4.2	Problématiques de discussion sur le modèle FP-péri	175

Figure 4.3	Exemple d'un objet de la catégorie 'animaux' utilisé dans l'Expé- rience 5	179
Figure 4.4	Deux types de contextes pouvant donner lieu à deux types de prédictions de 'bas niveau'.	180
Figure 4.5	Précision du modèle FP-péri	181
Figure 4.6	Stimuli et résultats de l'étude de Balas et al. (2009)	186
Figure 4.7	Stimuli de l'étude de Ehinger & Rosenholtz (2016)	187
Figure 4.8	Modèle de hiérarchie inverse	190

Liste des tableaux

Tableau 1.1	Paramètres expérimentaux des études étayant le modèle FP . . .	55
Tableau 2.1	Résultats de l'Expérience 1 (Session 1)	74
Tableau 2.2	Résultats de l'Expérience 1 (Session 2)	77
Tableau 2.3	Résultats de l'Expérience 2	92
Tableau 2.4	Résultats de l'Expérience 3	95
Tableau 2.5	Résultats de l'Expérience 4	107
Tableau 2.6	Résultats de l'Expérience 5	118
Tableau 3.1	Classification HPA (Hodapp, Parish & Anderson, 1993) de la sévérité de l'atteinte périmétrique glaucomateuse.	130
Tableau 3.2	Résultats de l'Expérience 6 - Session expérimentale SCENE . .	144
Tableau 3.3	Résultats de l'Expérience 6 - Session expérimentale VISAGE . .	147
Tableau 3.4	Caractéristiques et données cliniques des participants de l'Expé- rience 7	154
Tableau 4.1	Taille de l'effet d'interférence dans les expériences du Chapitre 2	170

Préambule

Les travaux présentés dans cette thèse ont été réalisés au sein du Laboratoire de Psychologie et Neurocognition (LPNC, CNRS UMR 5105) et de l'Ecole Doctorale d'Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE), grâce à un financement de l'Université Grenoble Alpes (Allocation de recherche doctorale AGIR POLE CBS).

Ces travaux s'inscrivent dans les thématiques de recherche de l'équipe Vision & Emotion du LPNC, et auparavant ceux de l'équipe Perception et Sensori-Motricité. Créée en janvier 2019, l'équipe Vision & Emotion développe des modèles de la perception visuelle en considérant tous ses aspects (traitement rétinien, vision active et mouvements oculaires, vision proactive, traitement émotionnel), et ce par le biais de mesures comportementales, physiologiques et neurophysiologiques et de simulations neurocomputationnelles. L'équipe Vision & Emotion travaille notamment depuis longtemps sur le développement et l'enrichissement de modèles théoriques de la reconnaissance visuelle de scènes. L'équipe s'est donné comme objectif général d'étudier les mécanismes qui permettent une reconnaissance visuelle rapide, automatique et efficace.

Dans ce contexte, ce travail de thèse s'est attaché à examiner quel pouvait être la contribution de la vision périphérique dans les mécanismes qui nous permettent de reconnaître rapidement notre environnement et les objets qui nous entourent. Plus spécifiquement, nous nous sommes demandés si la nature de l'information disponible dans le champ visuel périphérique pouvait influencer la manière dont nous percevons et catégorisons l'information perçue au niveau de la fixation, dans le champ visuel central.

Ces travaux ont été menés en collaboration avec l'équipe VIBS du laboratoire Grenoble Image Parole Signal Automatique (GIPSA-Lab, CNRS UMR 5216) pour les aspects de traitement du signal, l'Unité Mixte de Service IRMaGe (affiliée à l'INSERM US17, à l'Université Grenoble Alpes, au CHU Grenoble Alpes et au CNRS UMS 3552) pour l'accès à la plateforme IRM, et la clinique d'ophtalmologie du CHU Grenoble Alpes pour les études menées auprès de patients ophtalmologiques.

Chapitre 1

Cadre théorique : Modélisation de la reconnaissance visuelle

L'interprétation des signaux venant du monde extérieur est critique pour tout être vivant. L'évolution a façonné les systèmes perceptifs des espèces animales pour qu'elles puissent interpréter le monde extérieur le plus efficacement possible, afin de pouvoir produire les réponses appropriées. Chez l'humain, la vision est le sens dominant et celui qui apporte le plus d'informations sur le monde. Il est plutôt simple de se rendre compte de la rapidité avec laquelle nous pouvons prendre connaissance de notre environnement à l'aide de la vision par rapport à d'autres modalités sensorielles. Par exemple, si l'on attrape au hasard un objet, sans information préalable, quelques secondes peuvent être nécessaires pour estimer de quoi il pourrait s'agir, alors que reconnaître un objet visuellement est un phénomène quasi instantané. Le caractère rapide et fiable de la reconnaissance visuelle passionne la psychologie et les sciences de la vision depuis longtemps et constitue aussi le point d'ancrage de ce travail de doctorat. L'objectif de ce premier chapitre est de présenter les modèles dominants de la reconnaissance visuelle et d'en extraire les points communs, puis d'argumenter en quoi il est important de considérer les différences de traitement entre la vision centrale et la vision périphérique dans ces modèles. Nous présentons ensuite les questions auxquelles les travaux que nous rapportons dans ce manuscrit ont tenté de répondre et nous détaillons les méthodologies utilisées.

1. Reconnaissance visuelle ultra-rapide

1.1. Définitions : reconnaissance et catégorisation visuelle

D'un point de vue phénoménologique, la reconnaissance visuelle correspondrait au sentiment d'attribuer ou de pouvoir attribuer, grâce aux fonctions visuelles, un sens à ce qui nous entoure. En réalité, il est plutôt difficile d'avoir une idée claire de ce sentiment, puisque nous nous trouvons la plupart du temps dans des endroits identifiables, entourés d'objets auxquels nous pouvons associer une fonction ou une identité. Une autre façon de concevoir la reconnaissance visuelle serait de passer par le chemin inverse, c'est-à-dire d'imaginer ce qu'il se passe lorsqu'on ne peut pas du tout attribuer un sens à un percept visuel. L'image en Figure 1.1 est une photographie qui a circulé sur internet au début de l'année 2019, accompagnée de la légende *name one thing in this photo* (en français : essayez d'identifier quelque chose sur cette photo). Sur cette image, aucun objet n'est identifiable avec certitude et il est également difficile d'imaginer dans quel type d'environnement la photographie a été prise. On peut imaginer que cette sensation bizarre et un peu frustrante est le négatif de la sensation - fluide, non bizarre - de reconnaître notre environnement. D'ailleurs, cette sensation est certainement proche de celle des patients devenus agnosiques à la suite d'une lésion cérébrale, qui ne peuvent associer une signification à ce qu'ils perçoivent, alors même qu'il n'y a pas de cécité.

D'un point de vue théorique, la définition de la reconnaissance visuelle sur laquelle nous nous appuyons dans ce manuscrit est celle d'une correspondance entre un contenu perceptif et une représentation interne abstraite porteuse de sens, à un niveau conscient ou non. Cette conception repose sur une approche modulaire et symbolique de la cognition. Bien que cette conception générale soit discutable (voir par exemple [Glenberg & Kaschak 2003](#)), elle constitue un socle théorique généralement admis par les psychologues qui travaillent dans le domaine des sciences de la vision. Notons que la notion de reconnaissance visuelle est très proche de celle de catégorisation visuelle, puisque reconnaître un objet ou un environnement implique de pouvoir le classer dans une catégorie sémantique donnée. On peut définir la catégorisation comme un processus cognitif automatique qui regroupe les choses qui partagent certaines propriétés. Par exemple, la catégorie 'table' regroupe les objets constitués d'une surface dure et plane soutenue à distance du sol par quatre pieds. La reconnaissance visuelle serait donc le corollaire subjectif et accessible à la conscience, de la catégorisation. Par ailleurs, nous utiliserons souvent le terme de catégorisation pour nous référer à la méthode que nous utilisons pour étudier la reconnaissance visuelle dans les expériences décrites dans ce manuscrit de thèse. En effet, nous avons systématiquement utilisé des tâches de catégorisation : le participant doit décider, lorsqu'on lui présente un stimulus, à quelle catégorie il appartient. Il s'agit donc d'une catégorisation explicite, contrôlée, vers laquelle l'attention



Figure 1.1: Image naturelle constituée d'objets non identifiables, issue de @melip0ne. "Name one thing in this photo." 23 avril 2019, 3h45. Tweet.

est dirigée, et qui n'est donc pas parfaitement identique aux mécanismes de catégorisation implicites qui permettent la reconnaissance visuelle dans la vie de tous les jours. Néanmoins, les tâches de catégorisation constituent une méthode fiable pour mesurer la reconnaissance visuelle en laboratoire.

1.2. Évaluation de la reconnaissance visuelle ultra-rapide

Mis à part le cas des agnosies visuelles ou pour le genre d'image comme celle de la Figure 1.1, reconnaître visuellement notre environnement est donc un phénomène automatique, qui paraît instantané et ne nécessite aucun effort contrôlé. Malgré cette impression d'instantanéité, les processus qui permettent la reconnaissance visuelle impliquent un certain nombre de traitements, et donc une certaine durée, que des recherches empiriques ont cherché à quantifier. Une méthode d'investigation consiste à présenter successivement plusieurs stimuli visuels de manière très rapide (quelques dizaines de millisecondes) afin de simuler la succession de fixations rapides typiques de la vision naturelle. On demande ensuite à des sujets d'identifier un stimulus cible dans cette rafale d'images. Ce type de paradigme expérimental est souvent dénommé RSVP pour *Rapid Serial Visual Presentation*. En utilisant cette méthode, [Mary Potter \(1975\)](#) a démontré qu'il est possible d'interpréter une scène complexe lorsqu'elle est perçue pendant seulement 125 ms au sein d'une séquence RSVP. Les participants de cette étude devaient

identifier des scènes dans les séquences RSVP soit à partir de la présentation au préalable de la scène cible, soit à partir d'un label écrit attribué à la scène. Il n'est donc pas évident de déterminer ce qui a réellement pu être extrait de la scène en un temps si court dans cette expérience. En effet, les exemples fournis par l'auteure sont limités et on ne sait donc pas si les participants devaient principalement détecter des objets individuels (exemple donné par l'auteure : un bateau), plusieurs objets (deuxième exemple : un enfant et un papillon), ou une scène représentant des objets en interaction (troisième exemple : deux hommes buvant une bière).

Mais les travaux les plus cités à ce sujet, et qui permettent de mieux répondre à cette problématique, sont sans doute ceux de Michèle Fabre-Thorpe, Simon Thorpe et leur équipe. Dans l'étude princeps de [Thorpe, Fize, et Marlot \(1996\)](#), les auteurs présentaient à des participants des scènes naturelles qui contenaient un animal et d'autres qui n'en contenaient pas. Les scènes étaient flashées pendant seulement 20 ms et les participants devaient réaliser une tâche go/no-go. Ils devaient presser un bouton de réponse dès qu'ils percevaient un animal dans une scène (essais go) ou retenir leur réponse lorsqu'ils ne percevaient pas d'animal (essais no-go). Des réponses motrices correctes ont été enregistrées à partir de 300 ms seulement après le début de la présentation du stimulus. La mesure du temps de réponse est l'une des plus utilisées en psychologie expérimentale, et elle est très utile pour estimer le temps de traitement relatif de plusieurs processus. Cependant, cette mesure est nécessairement contaminée par le processus décisionnel et par le temps de programmation et d'exécution de la réponse motrice. C'est pourquoi les auteurs ont également mesuré les potentiels évoqués, à l'aide de l'électroencéphalographie (EEG), afin d'avoir une idée plus juste du temps nécessaire à la reconnaissance d'un animal dans la scène. L'analyse des potentiels évoqués visuels a ainsi révélé une négativité frontale associée aux essais no-go, mais pas aux essais go, 150 ms après le début de la présentation du stimulus. Les auteurs ont interprété cette différenciation des potentiels évoqués comme étant liée à des processus liés à la décision (e.g., l'inhibition de la réponse pour les essais no-go), qui ne peuvent avoir lieu qu'une fois que le stimulus est reconnu. Les résultats de cette étude suggèrent donc que seulement 150 ms suffiraient à analyser une scène et reconnaître une catégorie cible d'objet (ici un animal). Des délais très similaires (autour de 150 ms) ont été de nouveau observés dans une réplique de l'expérience, à la fois pour la détection d'animaux et pour la détection d'une nouvelle catégorie cible (les véhicules, [VanRullen & Thorpe 2001](#) ; voir aussi [Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe 2005](#)). Les travaux de cette même équipe de recherche ont aussi permis de montrer que les performances pour détecter des cibles dans des scènes naturelles sont aussi bonnes lorsqu'il s'agit de scènes déjà vues par les participants que lorsqu'il s'agit de nouvelles scènes ([Fabre-Thorpe, Delorme, Marlot, & Thorpe 2001](#)). Avec le même paradigme expérimental, [Delorme, Richard, et Fabre-Thorpe \(2000\)](#) ont comparé la détection d'aliments et d'animaux dans des scènes en couleurs (flashées pendant 32 ms) à la détection de ces cibles dans les mêmes scènes en niveaux de gris. De manière intéressante, leurs résultats ne montrent pas de différence de temps de réponse pour détecter des animaux que les images soient en couleurs ou en niveaux de gris. Pour

les aliments, la détection était plus rapide lorsque les images étaient en couleurs, mais la différence était petite (de l'ordre de 10 à 15 ms). Les auteurs ont alors proposé que les réponses de catégorisation rapide pourraient être basées sur un traitement minimal et grossier de la structure de la scène, sans analyse des indices de couleurs. Comme nous le verrons plus tard dans ce chapitre, ce type de traitement impliquerait des voies nerveuses magnocellulaires rapides. Cette particularité neurophysiologique est l'une des pierres angulaires des modèles de reconnaissance visuelle. Plus tard, [Joubert, Rousselet, Fize, et Fabre-Thorpe \(2007\)](#) ont montré que la catégorisation de scènes présentées pendant seulement 26 ms pouvait être aussi rapide que la détection d'objet. Les participants de cette expérience devaient catégoriser les scènes comme étant des scènes artificielles (i.e. façonnées par l'homme, comme par exemple une vue d'une pièce d'intérieur ou une vue de ville) ou des scènes naturelles (i.e. des paysages de nature comme un champ ou une plage). Les résultats montrent des temps de réponse médians autour de 390 ms. Les temps de réponse pour catégoriser des scènes à un niveau plus détaillé, c'est-à-dire suivant différentes catégories sous-ordonnées (mer, montagnes, intérieur et urbain), sont quant à eux plus longs, avec des temps de réponse médians de l'ordre de 400 à 460 ms selon la catégorie ([Rousselet, Joubert, & Fabre-Thorpe 2005](#)). Il s'agit d'un résultat important à considérer dans la suite de ce manuscrit, car il indique de nouveau que la première catégorisation pourrait être basée sur un traitement très grossier de la scène, et que cette catégorisation précéderait l'accès à des catégories plus précises.

Plus récemment, Simon Thorpe et ses collaborateurs ([Boucart et al. 2016](#); [Crouzet, Kirchner, & Thorpe 2010](#); [Kirchner & Thorpe 2006](#)) ont utilisé la latence des mouvements oculaires, plus rapides que les réponses manuelles, pour évaluer la vitesse de reconnaissance d'objets ou de visages dans une scène visuelle. Le principe expérimental est celui du paradigme de choix saccadique. Il consiste à flasher deux images de part et d'autre du point de fixation. Le participant a pour tâche de diriger son regard le plus rapidement possible du côté où l'image appartenant à une catégorie cible a été flashée. [Kirchner et Thorpe \(2006\)](#) ont observé que des saccades pouvaient être déclenchées du côté où se trouve un animal avec une bonne précision et en seulement 120-130 ms, les images ayant été flashées pendant seulement 20 ms. [Crouzet et al. \(2010\)](#) ont utilisé cette méthode pour évaluer la reconnaissance rapide de visages humains, d'animaux et de véhicules présentés pendant 400 ms. Les résultats ont montré que des saccades peuvent être déclenchées du côté où se trouve la catégorie cible avec une très bonne précision (environ 75% de saccades correctes pour les véhicules, 82% pour les animaux et 95% pour les visages) et avec des latences très courtes, les saccades les plus précoces de 110 ms ayant été observées pour les visages (des saccades correctes ont été initiées en seulement 120 ms pour les animaux, 140 ms pour les véhicules et 110 ms pour les visages). Plus récemment, les travaux de Nathalie Guyader et ses collaborateurs ([Guyader, Chauvin, Boucart, & Peyrin 2017](#); [Kauffmann et al. 2019](#)) ont de nouveau mis en évidence la rapidité des saccades oculaires dans le paradigme de choix saccadique (bien que le temps d'initiation des saccades soit légèrement plus long : 130 ms vers les visages, 170 ms vers les animaux et 180 ms vers les véhicules).

L'ensemble des travaux que nous venons d'exposer mettent en évidence la rapidité de la reconnaissance visuelle. En quelques dizaines de millisecondes, il est possible de se représenter le monde qui nous entoure et de catégoriser à la fois le contexte (la scène) et certains objets qui s'y trouvent. Les modèles de reconnaissance visuelle doivent donc absolument pouvoir rendre compte de cette rapidité de traitement. Comme nous l'avons rapidement évoqué avec les études de [Delorme et al. \(2000\)](#) et [Joubert et al. \(2007\)](#), la reconnaissance visuelle rapide (i.e. la reconnaissance possible lors des premiers instants de la fixation oculaire) pourrait être basée sur des représentations minimales et grossières de l'environnement et des objets qui s'y trouvent, plutôt que sur des représentations précises et détaillées. L'idée d'une première représentation rudimentaire accessible très rapidement et qui contourne le traitement des détails est centrale dans ce travail de thèse. Une telle perception de la scène visuelle est souvent qualifiée de *gist perception*, une expression que l'on pourrait traduire par perception de l'essentiel, ou perception globale. Nous utiliserons cependant le plus souvent le terme anglophone *gist*, parce qu'il est simple et fréquemment utilisé dans la littérature.

1.3. La perception de l'essentiel : le concept de *gist* de la scène

Dans la fin des années 1990 et le début des années 2000, Aude Oliva a consacré une grande partie de ses recherches à caractériser la reconnaissance rapide de stimuli naturels. En particulier, elle a proposé une conceptualisation du *gist* d'une scène visuelle. On peut résumer sa définition du *gist* comme ceci : **(i)** le *gist* correspond à ce qui peut être saisi d'une scène visuelle complexe en un seul coup d'œil et **(ii)** il correspond à l'expérience phénoménologique d'acquérir une compréhension de toute la scène d'un seul coup, indépendamment de son niveau de détails et d'encombrement ([Oliva 2005](#)). La nature exacte du *gist*, autrement dit à quel niveau il est représenté, est néanmoins plus complexe à déterminer. Oliva en propose une vision assez large, qui considère à la fois les propriétés de bas niveau des scènes (*gist* perceptuel) et les représentations abstraites/sémantiques qui peuvent en être extraites (*gist* conceptuel). Nous allons détailler, dans les paragraphes qui suivent, ces deux aspects des représentations *gist*.

1.3.1. Contenu perceptuel du *gist* : rôle des basses fréquences spatiales et des *summary statistics*

Selon [Oliva \(2005\)](#), le *gist* perceptuel fait référence à la représentation structurelle de la scène. Autrement dit, les propriétés perceptives du *gist* décrivent la structure, ou configuration spatiale, de la scène (i.e. la présence de différents éléments et leurs interrelations spatiales). Par exemple, la configuration spatiale d'une scène de plage peut être décrite en agençant trois larges bandes horizontales, de haut en bas : une pour le ciel, une plus fine pour la mer et une pour la plage de sable. L'accès à une représentation de la configuration spatiale de la scène nécessite une analyse grossière, à grande échelle spatiale mais ne requiert pas d'analyser les détails contenus dans les différentes régions

structurantes de la scène. Comme nous le verrons un peu plus loin dans ce chapitre, les informations extraites à grande échelle spatiale sont contenues dans les basses fréquences spatiales (BFs) du signal visuel, c'est-à-dire dans les larges variations de luminance. Au contraire, les détails, les contours précis des objets dans la scène, qui permettent une analyse plus fine, sont contenus dans les hautes fréquences spatiales (HFs) du signal visuel. Si, comme le suggère Oliva, l'extraction du *gist* perceptuel permet d'avoir accès rapidement à la configuration spatiale de la scène, alors les BFs, qui permettent de décrire cette configuration spatiale, devraient être disponibles très rapidement lors de l'analyse visuelle. Avec leur expérience notoire réalisée avec des images hybrides, Schyns et Oliva (1994) ont apporté une importante contribution à l'hypothèse de l'importance des BFs dans la reconnaissance visuelle rapide. Les images hybrides sont construites en superposant une scène filtrée qui ne préserve que les BFs à une autre scène filtrée qui ne préserve que les HFs. Dans cette expérience, les deux scènes appartenaient à deux catégories sémantiques différentes. Par exemple, une image hybride pouvait contenir les BFs d'une scène d'autoroute superposées aux HFs d'une scène de ville (Figure 1.2). L'image hybride était présentée brièvement à des participants, suivie d'une scène cible non filtrée. Dans la moitié des essais, la scène cible était en fait la version non filtrée du composant BFs ou du composant HFs de l'hybride. Dans l'autre moitié des essais, la scène cible était une nouvelle scène (i.e. elle n'était appariée ni à la scène en BFs ni à la scène en HFs de l'hybride). La tâche des participants était d'indiquer si la scène cible non filtrée était présente dans l'image hybride. Lorsque l'image hybride était présentée pendant 30 ms, les participants avaient plus tendance à dire que la scène cible était présente dans l'hybride lorsqu'elle correspondait à l'image en BFs de l'hybride que lorsqu'elle correspondait à l'image en HFs. Lorsque l'image hybride était présentée pendant 150 ms, le pattern inverse était observé : les participants disaient plus souvent que la scène cible était présente dans l'hybride lorsqu'elle correspondait avec les HFs que lorsqu'elle correspondait avec les BFs. À noter qu'une condition contrôle permettait aux auteurs de s'assurer qu'à la fois les scènes en BFs et en HFs de l'hybride pouvaient être correctement perçues à 30 ms comme à 150 ms de présentation. Les résultats de cette étude indiquent que le traitement des fréquences spatiales évolue au cours du temps. Les BFs domineraient la perception lors du traitement rapide des scènes alors que l'analyse de l'information fine véhiculée par les HFs serait privilégiée lorsque plus de temps est alloué au traitement de la scène. La rapidité de traitement des BFs relativement au traitement des HFs a été observée dans de nombreuses études sur lesquelles nous reviendrons. L'important à retenir ici est que les BFs d'une scène visuelle seraient disponibles très tôt et pourraient donc être l'essence même, en termes de propriétés perceptives, du *gist* de la scène.

De notre point de vue, les BFs ne sont pas nécessairement les seules propriétés perceptives représentées dans le *gist*. Notre environnement visuel contient de nombreuses surfaces plutôt uniformes (e.g., le ciel, une façade d'immeuble), mais aussi des régions où se répètent les mêmes motifs (e.g., les feuilles d'un arbre, des voitures alignées sur un parking, un mur de briques). Aussi, le signal visuel est généralement redondant dans des régions proches du champ visuel, et deux points proches ont donc de grandes chances

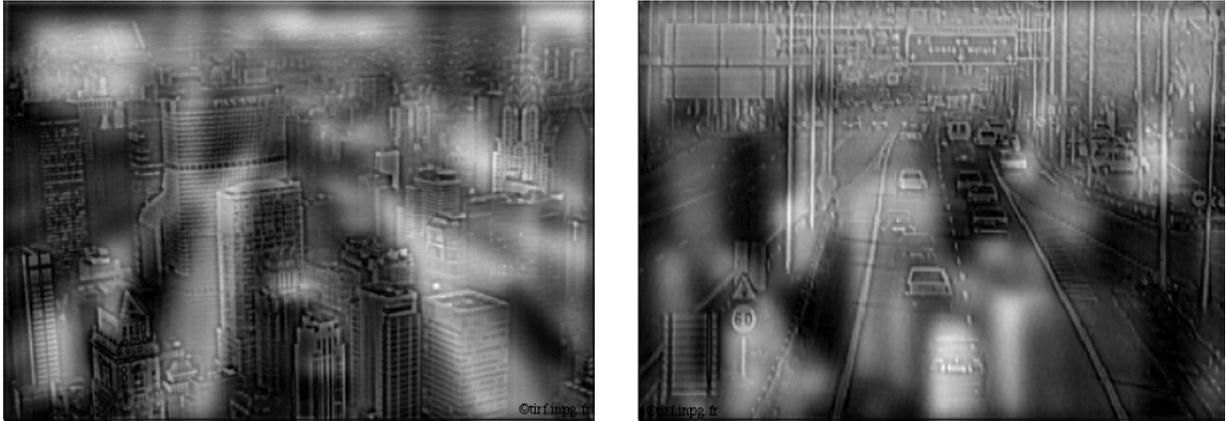


Figure 1.2: Exemples d'images hybrides. L'image hybride de gauche représente une scène de ville dans les HFs et une autoroute dans les BFs. L'image hybride de droite est complémentaire. Elle représente une autoroute dans les HFs et une scène de ville dans les BFs. Tiré de [Schyns et Oliva \(1994\)](#).

de corrélérer sur de nombreuses dimensions (e.g., couleur, luminance, texture, orientation, mouvement, motif, etc.). Le concept de *summary statistics* (ou perception d'ensemble) fait référence à la capacité du système visuel à calculer des estimateurs moyens de ces diverses propriétés de la scène visuelle ([Whitney & Yamanashi Leib 2018](#)). Ces mécanismes sont économiques computationnellement, puisque le système visuel ne représenterait pas chaque élément d'une scène de manière individuelle (e.g., chaque feuille de l'arbre ou chaque brique du mur), mais regrouperait plutôt les éléments similaires pour en extraire la distribution (e.g., l'orientation moyenne des feuilles, ou la taille moyenne des briques ; [Alvarez 2011](#)). Les *summary statistics* peuvent, à notre sens, être également considérés comme une composante du *gist* perceptuel.

Une des premières démonstrations expérimentales du phénomène de *summary statistics* est celle d'[Ariely \(2001\)](#). Dans cette expérience, chaque essai consistait en la présentation d'un pattern de disques blancs de différents diamètres dans un premier intervalle, suivi de la présentation d'un seul disque (disque test) dans le second intervalle (Figure 1.3). La tâche des participants était de décider si le disque test était présent parmi le pattern de disques présenté dans le premier intervalle, ou non. Les résultats ont montré que la probabilité de juger que le disque test fait partie du pattern est plus grande lorsque le diamètre du disque test est proche du diamètre moyen de l'ensemble des cercles. Plutôt que d'encoder la taille de chacun des cercles, il serait donc plus économique de représenter la taille moyenne de l'ensemble des cercles. Le pattern de disques et le disque test étaient tous deux présentés pendant 500 ms, indiquant que ce laps de temps suffirait à pouvoir s'appuyer sur ces mécanismes de *summary statistics*.

[Chong et Treisman \(2003\)](#) ont présenté à des participants deux ensembles de cercles de diamètres hétérogènes pendant 100 ms seulement, l'un à gauche de l'écran et l'autre à droite. Ils devaient juger lequel parmi les deux ensembles avait le diamètre moyen

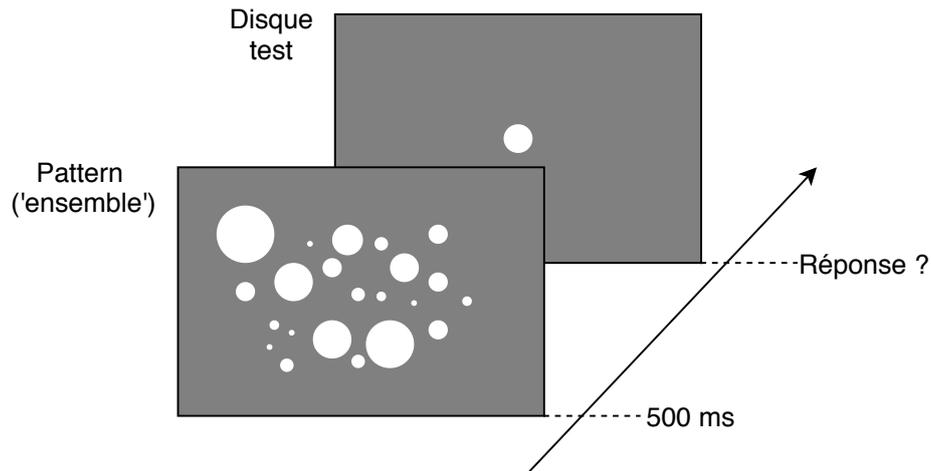


Figure 1.3: Représentation schématique de la procédure expérimentale et des stimuli utilisés par Ariely (2001).

le plus grand. Les résultats ont montré qu'il suffisait que les diamètres moyens des deux ensembles diffèrent seulement d'environ 7% pour que les participants atteignent 75% de réponses correctes, mettant en exergue la précision avec laquelle les *summary statistics* peuvent être calculés. Une deuxième expérience montrait qu'un tel pattern de résultats peut être également observé avec seulement 50 ms de présentation. La capacité à pouvoir estimer des tailles moyennes serait donc remarquablement précise et rapide. La possibilité de calculer des *summary statistics* en des temps très courts a également été montrée pour d'autres caractéristiques visuelles de bas niveau. Par exemple, il est possible d'estimer l'orientation moyenne de patches de Gabor présentés pendant 100 ms (Parkes, Lund, Angelucci, Solomon, & Morgan 2001) ou la luminance moyenne d'un ensemble de disques présenté pendant 500 ms (Bauer 2009).

Des *summary statistics* peuvent aussi être extraits de stimuli plus complexes, comme un ensemble de visages. Par exemple, avec un paradigme expérimental similaire à celui d'Ariely (2001), de Fockert et Wolfenstein (2009) ont demandé à des participants de juger si un visage était présent ou non dans un ensemble de visages présenté au préalable. Le visage test pouvait être soit un membre de l'ensemble, soit un non-membre, soit un morphing de l'ensemble des visages (morphing apparié), soit un morphing d'autres visages (morphing non apparié; Figure 1.4). Leurs résultats montrent que la probabilité de juger que le visage test fait partie de l'ensemble est plus grande lorsque celui-ci est un membre ou un morphing apparié que lorsque c'est un non membre ou un morphing non apparié. Cette probabilité était même plus grande pour le morphing apparié que pour le visage membre. Les identités d'un ensemble de visages pourraient alors être représentées avec une description moyenne. Par ailleurs, Haberman et Whitney (2009) ont montré que des participants étaient capables d'extraire "l'émotion moyenne" d'un ensemble de visages présenté pendant seulement 50 ms, suggérant qu'il est également possible de saisir rapidement l'état émotionnel moyen d'un groupe de personnes.

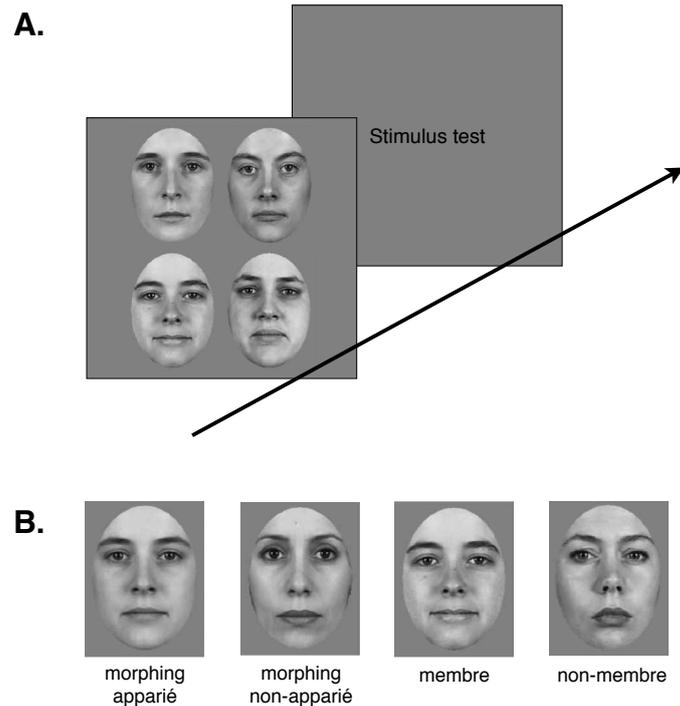


Figure 1.4: **A.** Procédure expérimentale de l'expérience de [de Fockert et Wolfenstein \(2009\)](#) : un ensemble de visage apparaît, puis le stimulus test. **B.** Le stimulus test pouvait être un membre de l'ensemble de visage, un non-membre, un morphing de l'ensemble des visages (morphing apparié), ou un morphing d'autres visages (morphing non apparié).

Pour résumer, le système visuel serait capable d'extraire de manière quasi immédiate certains contenus visuels qui forment le *gist* perceptuel. Notamment, les travaux que nous avons exposé indiquent que la configuration spatiale de la scène pourrait être disponible rapidement grâce à une analyse très grossière du signal de BFs. Le système visuel pourrait également s'appuyer sur des *summary statistics* qui permettent d'estimer le "prototype" de diverses propriétés visuelles de la scène, comme des orientations, ou mêmes des indices socialement pertinents comme l'émotion moyenne d'une foule.

1.3.2. Contenu conceptuel du *gist* : Quel niveau de catégorisation au premier coup d'œil ?

Le *gist* conceptuel fait référence aux contenus sémantiques qui peuvent émerger rapidement lors de la perception d'une scène visuelle. Dans les années 1970, les travaux d'Elisabeth Rosch sur l'organisation des connaissances conceptuelles ont montré que le niveau de base serait le niveau de catégorisation favorisé pour catégoriser les objets et le premier à être accessible. Le niveau de base est un niveau de spécificité intermédiaire. Il correspond au niveau de catégorisation permettant de faire le meilleur compromis entre la maximisation de la ressemblance intra-catégorielle d'une part et de la variabilité inter-catégorielle d'autre part ([Mervis & Rosch 1981](#); [Rosch & Lloyd 1978](#); [Rosch, Mervis,](#)

Gray, Johnson, & Boyes-Braem 1976). La catégorisation d'objets au niveau de base (e.g., c'est une voiture) serait donc favorisée par rapport à une catégorisation au niveau sur-ordonné (e.g., c'est un moyen de transport) ou sous-ordonné (e.g., c'est une Fiat Uno). Il s'agit également du niveau choisi par les enfants pour dénommer des images d'objets (Rosch et al. 1976). Ce principe est d'ailleurs souvent observé dans le langage courant (R. Brown 1958). Par exemple, on dira plus souvent "je suis assise sur une chaise, je mange une pomme" (niveau de base), plutôt que "je suis assise sur une chaise de bureau, je mange une reinette du Canada" (niveau de catégorisation sous-ordonnée) ou encore "je suis assise sur un meuble, je mange un fruit" (niveau de catégorisation sur-ordonnée). On peut alors se demander si le niveau de base est favorisé dès les premiers instants de la reconnaissance visuelle.

L'étude de Guyader et al. (2017) montre que les saccades oculaires vers les visages humains, lorsque les distracteurs sont des véhicules ou des animaux, sont extrêmement rapides. De façon intéressante, cet avantage est accentué lorsque seul le contenu en BFs est disponible, indiquant que cette information rudimentaire serait suffisante pour différencier très rapidement un visage d'un véhicule ou d'un animal, ce qui peut s'apparenter à une catégorisation au niveau sur-ordonné. À noter toutefois que les visages représentent une catégorie d'objets bien particulière chez l'humain en raison de l'exposition fréquente à ces stimuli et surtout de leur importance sociale. Les études en neuroimagerie montrent qu'il y a des régions du cortex occipito-temporal qui seraient spécifiquement impliquées dans le traitement de ce genre de stimuli, notamment une région du gyrus fusiforme (*Fusiform Face Area*; Kanwisher, McDermott, & Chun 1997). En outre, des travaux suggèrent que le traitement holistique des visages reposerait sur l'analyse des BFs (Goffaux & Rossion 2006). La catégorisation rapide au niveau sur-ordonné dans l'étude de Guyader et al. (2017) pourrait donc simplement s'expliquer par une spécificité du traitement des visages et par un rôle important des BFs dans celui-ci. Mais même au-delà du cas particulier des visages, les expériences de catégorisation visuelle rapide, utilisant par exemple des paradigmes en go/no-go, ont remis en question l'hypothèse de la primauté du niveau de base de Rosch. Nous avons vu plus haut que reconnaître des objets dans des scènes au niveau sur-ordonné est extrêmement rapide (e.g., détecter un animal, un véhicule, ou encore un aliment - Bacon-Macé et al. 2005; Delorme et al. 2000; Thorpe et al. 1996; VanRullen & Thorpe 2001). D'autres études ont directement étudié la rapidité de la reconnaissance au niveau de base *vs* au niveau sur-ordonné, et montrent que les réponses pour reconnaître les objets au niveau sur-ordonné sont plus rapides et plus justes (Macé, Joubert, Nespoulous, & Fabre-Thorpe 2009; Ramezani, Kheradpisheh, Thorpe, & Ghodrati 2019; Wu, Crouzet, Thorpe, & Fabre-Thorpe 2015). Lors de la reconnaissance visuelle rapide, des représentations au niveau sur-ordonné émergeraient alors plus rapidement que des représentations au niveau de base. Macé et al. (2009) font d'ailleurs justement remarquer que la plupart des études mettant en avant une primauté du niveau de base demandent une réponse verbale, alors qu'un tel traitement lexical n'est pas nécessaire dans les paradigmes de catégorisation visuelle rapide.

Concernant la catégorisation de scènes, certains auteurs ont là aussi postulé que le niveau de base serait favorisé (e.g., montagne, plage, autoroute; [Tversky & Hemenway 1983](#)). À l'appui de cette hypothèse, des travaux ont montré que différents exemplaires de scènes de la même catégorie niveau de base tendent à avoir la même configuration spatiale, les mêmes couleurs dominantes, mais aussi une distribution similaire de l'amplitude sur les différentes orientations et fréquences spatiales (i.e. une correspondance des spectres d'amplitude; [Torralba & Oliva 2003](#)). Comme nous l'avons vu plus haut, les scènes peuvent effectivement être catégorisées au niveau de base alors même que leur présentation est très brève ([Joubert et al. 2007](#), scènes flashées pendant seulement 26 ms). Grâce à un paradigme de masquage, où la présentation d'un masque suivait la présentation de la scène à catégoriser avec un intervalle variable, [Greene et Oliva \(2009a\)](#) ont montré que la catégorisation de scènes au niveau de base atteignait 75% de précision lorsque les scènes étaient masquées 50 ms après le début de leur présentation. Un temps de présentation aussi court que 50 ms pourrait ainsi suffire à accéder à la catégorie niveau de base d'une scène. Cependant, dans la même étude, ces auteures utilisaient aussi une autre tâche qui consistait à catégoriser les scènes non pas en fonction de leur niveau de base, mais en fonction de leurs "propriétés globales". Cette typologie proposée par les auteures correspond plutôt aux propriétés fonctionnelles et structurelles de la scène, telles que sa profondeur, sa navigabilité, son degré d'ouverture, ou son caractère naturel (*vs* urbain). Les résultats de cette tâche ont montré que le délai de masquage nécessaire pour atteindre 75% de précision était significativement plus court que pour la catégorisation au niveau de base, à savoir 34 ms. La reconnaissance des propriétés globales des scènes pourrait alors précéder la catégorisation au niveau de base. De manière intéressante, dans une autre étude ([Greene & Oliva 2009b](#)), ces mêmes auteures proposent que la catégorisation sémantique au niveau de base ne constitue pas nécessairement l'essence de la reconnaissance visuelle rapide, mais serait plutôt une conséquence de représentations basées sur l'analyse des propriétés globales des scènes. En effet, les participants de cette étude, qui devaient identifier des scènes appartenant à une catégorie de niveau de base cible parmi des distracteurs, faisaient davantage de réponses de type fausse alarme pour des distracteurs dont les propriétés globales étaient similaires à celles prototypiques de la catégorie cible. Par exemple, lorsque les participants devaient identifier des scènes de forêts, qui ont un faible niveau d'ouverture, les scènes distractrices à faible niveau d'ouverture provoquaient des fausses alarmes. Similairement, lorsqu'ils devaient identifier des scènes de champs, qui ont un haut niveau d'ouverture, c'était cette fois-ci les scènes distractrices à haut niveau d'ouverture qui provoquaient des fausses alarmes. Par ailleurs, rappelons que nous avons remarqué que les temps de réponse pour catégoriser des scènes à un niveau sur-ordonné (scènes artificielles *vs* scènes naturelles, [Joubert et al. 2007](#)) étaient plus courts (390 ms) que ne l'étaient les temps de réponse pour catégoriser des scènes au niveau de base ([Rousselet et al. 2005](#), 400 à 500 ms selon les catégories). Dans un paradigme de masquage similaire à celui utilisé par [Greene et Oliva \(2009a\)](#), [Loschky et Larson \(2010\)](#) ont comparé explicitement une catégorisation au niveau de base (plage, montagne, ville, autoroute) à une catégorisation sur-ordonnée (artificiel *vs*

naturel). Leurs résultats ont montré qu'avec des temps de présentation allant jusqu'à environ 70 ms, la précision de catégorisation au niveau sur-ordonné était meilleure que celle au niveau de base (pour des résultats similaires, voir aussi [Kadar & Ben-Shahar 2012](#); [Renninger & Malik 2004](#)). Par ailleurs, une étude récente a montré que la catégorisation rapide de scènes au niveau sur-ordonné serait moins dérangée par une occupation simultanée de la mémoire de travail relativement à la catégorisation rapide de scènes au niveau de base ([Sun, Ren, Zheng, Sun, & Zheng 2016](#)).

Lors des premières dizaines de millisecondes, une catégorisation sur-ordonnée basée sur les propriétés globales des scènes pourrait donc précéder une catégorisation au niveau de base de quelques dizaines de millisecondes. Cette hypothèse reste toutefois à modérer. [Fei-Fei, Iyer, Koch, et Perona \(2007\)](#) ont cherché à déterminer ce qui pouvait être saisi d'une scène visuelle en quelques dizaines de millisecondes en utilisant un paradigme de report subjectif. Les auteurs présentaient des scènes visuelles très variables, pendant des temps de présentations brefs (de 27 à 500 ms). Les participants décrivaient librement ce qu'ils avaient pu percevoir de la scène. Par exemple, un participant décrivit une scène présentée pendant 67 ms comme ceci : "Ça ressemblait à une scène d'intérieur. J'ai vu ce qui ressemblait à un grand objet encadré (une peinture ?) sur un fond blanc (sûrement le mur)". Afin de déterminer quel type de contenu sémantique pouvait être saisi lors de la perception du *gist* des scènes, un groupe de participants qui n'avaient pas réalisé l'expérience évaluait ensuite le contenu des descriptions en les classant en fonction de différents niveaux de catégorisation (i.e. niveau sur-ordonné, niveau de base, niveau sous-ordonné). Contrairement à ce qui a été observé avec des paradigmes de catégorisation ([Greene & Oliva 2009b](#); [Loschky & Larson 2010](#)), les résultats de cette étude n'ont mis en évidence aucune préférence d'un niveau de catégorisation par rapport à un autre, ni pour la description des scènes dans leur globalité (e.g., pour un temps de présentation donné, les participants pouvaient aussi bien décrire une scène comme étant une scène d'intérieur que comme étant un salon), ni pour la description d'objets (e.g., un objet pouvait être décrit aussi bien comme un meuble que comme une table). Cette méthode de report subjectif pourrait alors suggérer que les catégories accessibles lors des premiers instants de la reconnaissance visuelle sont variables et qu'il est difficile d'établir une hiérarchie des traitements. Toutefois, il faut garder en tête qu'ici encore une réponse verbale est exigée du participant, et que les réponses motrices ou oculomotrices permettent sûrement une meilleure mesure de la reconnaissance visuelle rapide puisque non contaminées par les processus lexicaux.

Dans cette partie, nous avons défini le *gist* d'une scène visuelle comme étant une représentation rudimentaire de la scène, accessible très rapidement. Pour résumer, le *gist* pourrait contenir une information sur la structure de la scène contenue dans les BFs, ainsi que des *summary statistics* qui permettent de décrire très rapidement diverses propriétés de la scène. Ces contenus perceptifs permettraient de saisir les aspects conceptuels / sémantiques du *gist*, et ainsi de catégoriser rapidement les scènes, et éventuellement quelques objets dans la scène, plutôt à un niveau sur-ordonné. Nous verrons dans la

prochaine section que l'idée d'une perception du *gist* des scènes visuelles est centrale dans les modèles de reconnaissance visuelle.

2. Du modèle fréquentiel à un modèle fréquentiel prédictif de la reconnaissance visuelle

L'objectif de cette partie est de faire une présentation de différents résultats empiriques et propositions théoriques visant à expliquer la perception visuelle, et notamment de deux modèles de reconnaissance visuelle complémentaires qui ont été développés depuis les années 2000. Nous décrivons d'abord le modèle fréquentiel (Marendaz, Rousset, & Charnallet 2003), puis le modèle proactif de Moshe Bar (2003, 2007, Bar et al. 2006). Nous allons voir que ces modèles s'appuient, explicitement ou non, sur l'idée d'une extraction rapide du *gist* de la scène. Parmi toutes les propriétés du *gist* que nous avons décrites, ces modèles s'intéressent particulièrement à l'information issue du traitement rapide des BFs. En effet, ils sont basés sur l'hypothèse d'un principe de traitement par défaut *coarse-to-fine* de l'information visuelle. Selon cette hypothèse, le traitement des BFs du signal visuel précéderait celui des HF. Le modèle fréquentiel postule que ce principe gouverne les mécanismes de reconnaissance visuelle, et le modèle proactif postule plus spécifiquement que les premiers traitements grossiers - reposant sur l'analyse des BFs - permettraient d'activer des mécanismes de prédictions, lesquels seraient utilisés pour faciliter l'analyse subséquente des détails. Avant de développer ces deux modèles de reconnaissance visuelle, nous définirons ce que sont les fréquences spatiales et nous verrons comment cette caractéristique élémentaire du signal visuel peut être manipulée, avec la technique du filtrage spatial, afin de tester expérimentalement le rôle des informations contenues dans les différentes bandes de fréquences spatiales. Nous verrons aussi que notre système visuel est particulièrement bien équipé pour traiter le signal visuel en fonction des différentes fréquences spatiales.

2.1. Le système visuel : un analyseur de Fourier

2.1.1. Les fréquences spatiales

En termes d'analyse du signal, la fréquence spatiale permet de définir le nombre d'ondulations de luminance par unité de distance. Elle est usuellement exprimée par le nombre de cycles (i.e. d'ondulations de luminance) par degré d'angle visuel (cpd) ou par le nombre de cycles par image (cpi). Le nombre de cycles par degré d'angle visuel est une mesure relative, qui varie en fonction de la distance entre l'observateur et le stimulus. Au contraire, le nombre de cycles par image est une mesure absolue, indépendante de l'observateur. Plus il y a de cycles par degré/par image, plus la fréquence spatiale est haute, et inversement (Figure 1.5). Les basses et hautes fréquences spatiales fournissent des informations différentes sur la scène visuelle. Les BFs correspondent à des 'blobs' (formes floues) de luminance. Comme déjà mentionné, ces larges blobs

de luminance permettent de véhiculer une information grossière, comme la structure globale de la scène et une représentation imprécise des formes et des éléments qui la composent. Les BF_s préservent les relations spatiales entre les éléments dans la scène, mais ne permettent pas la perception des détails et l'identification de petits objets. Les HF_s contiennent au contraire une information fine, détaillée, comme les contours précis des objets et les petits détails de la scène. C'est l'information en HF_s qui permettra une analyse localisée minutieuse et précise des régions de la scène. Si l'on prend l'exemple d'une photographie représentant une vue de la ville de Grenoble (Figure 1.5) filtrée pour préserver soit seulement les BF_s (filtre passe-bas) soit seulement les HF_s (filtre passe-haut), on voit que l'information concernant la structure de la scène et les formes principales qui la composent comme le pont, la rivière, la chaîne de montagnes, est contenue dans les BF_s, alors que l'information détaillée, comme les immeubles et leurs nombreuses fenêtres, le clocher de la cathédrale et les voitures, est contenue dans les HF_s. Ainsi, notre environnement visuel nous fournit une information riche qui se décline sur un large spectre spatial, d'une échelle spatiale très large (très basses fréquences spatiales) à une échelle spatiale très fine (très hautes fréquences spatiales). Ce spectre d'information peut être décrit en utilisant une analyse de Fourier.

2.1.2. Un outil pour étudier le rôle des fréquences spatiales dans la perception visuelle : Le filtrage spatial

La transformée de Fourier est une série d'opérations mathématiques qui permet de décomposer un signal visuel en une somme finie de sinusoides, dont chacune est caractérisée sur quatre dimensions : sa fréquence, son amplitude (contraste), sa phase et son orientation. Grâce à la transformée de Fourier, il est alors possible de décomposer une image et de la décrire par son spectre d'amplitude et son spectre de phase. Le spectre d'amplitude donne la distribution du contraste de luminance (i.e. l'amplitude) en fonction de l'orientation et de la fréquence spatiale. Le spectre de phase caractérise la position relative de la luminance sur les différentes fréquences spatiales dans l'espace de l'image. Sur le spectre d'amplitude, les BF_s de l'image sont représentées au centre tandis que les HF_s sont représentées en périphérie (Figure 1.6). À noter que pour les images naturelles, l'amplitude du signal décroît de manière inverse à la fréquence spatiale (i.e. le contraste est bien plus important dans les BF_s de l'image que dans les HF_s; Field 1987). Le facteur de décroissance correspond à peu près à une fonction $\frac{1}{f^\alpha}$, où f est la fréquence spatiale et α est un facteur variant entre environ 0.7 et 1.5 (Bex & Makous 2002).

En retirant certaines parties du spectre d'amplitude, il est donc possible de filtrer des bandes choisies de fréquences spatiales de l'image. En appliquant ensuite une transformée de Fourier inverse, on obtient alors des images dont le contenu en fréquence spatiale est contrôlé, ce qui fait de la transformée de Fourier un outil particulièrement utile pour étudier expérimentalement le rôle des fréquences spatiales dans la perception visuelle. Ainsi, si on applique un filtre passe-bas au spectre d'amplitude, l'image résultante après

Basses fréquences spatiales (BFs)

fréquences spatiales < 4 cpi



fréquences spatiales < 8 cpi



Hautes fréquences spatiales (HFs)

fréquences spatiales > 60 cpi

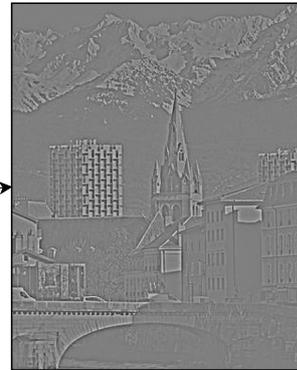
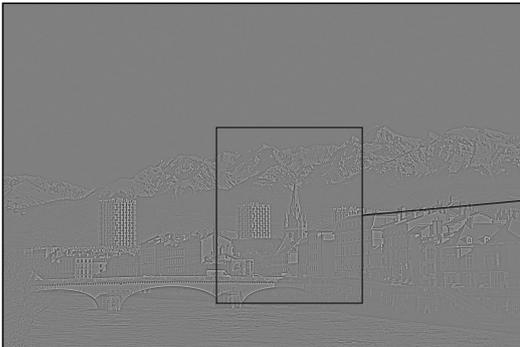


Figure 1.5: Illustration de l'information contenue dans les basses et hautes fréquences spatiales d'une scène naturelle. Des fréquences spatiales inférieures à 4 cycles par image (cpi), on ne peut extraire que la configuration spatiale de la scène (structure horizontale, grand degré d'ouverture). Les fréquences spatiales inférieures à 8 cpi permettent d'identifier certaines parties structurantes de la scène : la chaîne de montagne, la rangée d'immeubles, la rivière, le pont. Les très hautes fréquences spatiales (supérieures à 60 cpi) permettent au contraire de percevoir les détails : fenêtres, voitures, etc.

transformée de Fourier inverse sera composée seulement de BFs, et toutes les fréquences supérieures à une fréquence donnée (la fréquence de coupure) seront retirées de l'image. À l'inverse, un filtre passe-haut conservera seulement les fréquences spatiales supérieures à la fréquence de coupure choisie. À noter que le spectre de phase n'est pas altéré lors de ces manipulations, afin que la position relative des fréquences spatiales (donc la configuration spatiale de l'image) reste intacte.

2.1.3. Bases biologiques de l'analyse fréquentielle de l'information visuelle : Les systèmes magnocellulaire et parvocellulaire

Les modèles de la reconnaissance visuelle que nous allons décrire reposent sur l'hypothèse que le système visuel se comporte comme un analyseur de Fourier, c'est-à-dire

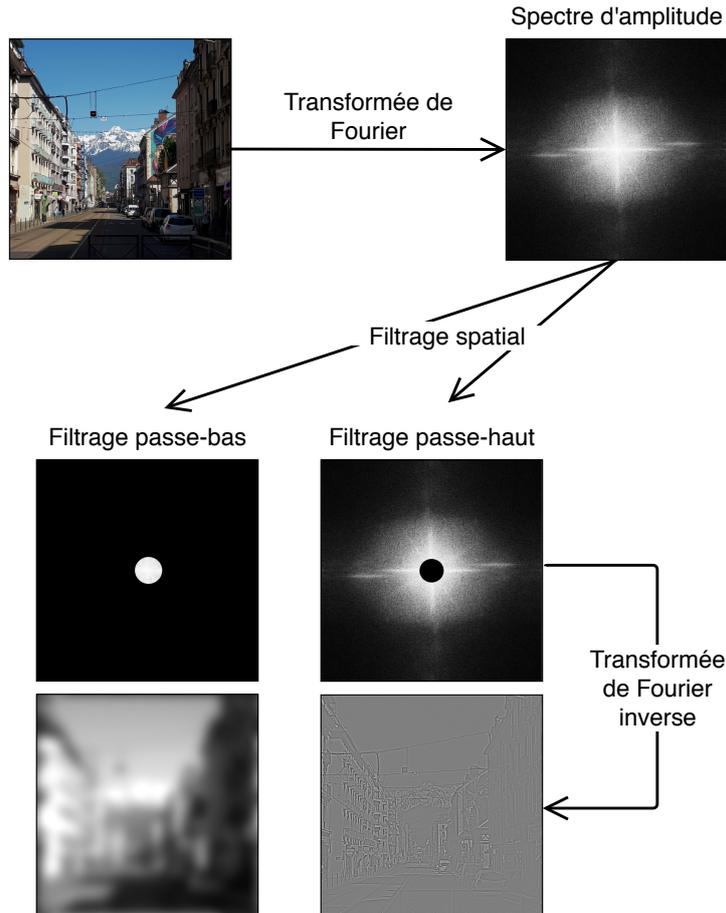


Figure 1.6: Illustration de la procédure de filtrage spatial d'une image

qu'il extraierait en parallèle les variations de luminance à différentes échelles spatiales et en fonction de différentes orientations. Cette hypothèse est cohérente avec certaines propriétés neurophysiologiques du système visuel, notamment les traitements distincts et en parallèle des basses et hautes fréquences spatiales par les cellules rétiniennes et les voies visuelles.

Lorsque la lumière pénètre dans l'œil, elle est d'abord absorbée par les cellules photoréceptrices qui convertissent les radiations électromagnétiques de la lumière en signal nerveux. On distingue deux types de cellules photoréceptrices sur la base de leurs propriétés anatomiques et physiologiques, les cônes et les bâtonnets. Les cônes sont sensibles aux hautes intensités lumineuses. Ils sont donc particulièrement actifs en condition de fort éclairage (vision photopique). Les bâtonnets peuvent au contraire répondre à de faibles intensités lumineuses et permettent donc la vision nocturne (scotopique). Au sein de la rétine, le signal électrique provenant des photorécepteurs est ensuite transmis aux cellules ganglionnaires par le biais des cellules bipolaires. Parmi les cellules ganglionnaires, on distingue les cellules dites parasols ou magnocellulaires (M), les cellules dites naines ou parvocellulaires (P), et les cellules koniocellulaires¹. Les cellules M ont

1. Les cellules koniocellulaires sont relativement mal connues et sont d'intérêt moindre pour ce travail de thèse. Elles ne seront donc plus abordées dans la suite du manuscrit.

de grands corps cellulaires et une arborisation dendritique étendue, ce qui implique que leur activité est modifiée par un grand nombre de photorécepteurs, donc par une information lumineuse provenant d'une relativement grande région de l'espace. On parle de cellules à grands champs récepteurs. Il faut souligner que les axones des cellules M sont larges et myélinisés et permettent ainsi une transmission rapide de l'information. Ces cellules répondent à des stimulations brèves et transitoires (réponse phasique) et sont sensibles aux variations de contraste achromatique et aux mouvements. Par contre, elles répondent seulement à des fréquences spatiales relativement basses, du fait de leurs larges champs récepteurs. Les cellules P ont au contraire un petit corps cellulaire et une arborisation dendritique restreinte, ce qui implique que leur activité est modifiée par un petit nombre de photorécepteurs. Au centre de la rétine, chaque cellule P reçoit même le signal provenant d'un seul cône, permettant ainsi une excellente résolution spatiale (Kaplan 2004). Ces cellules ont alors de petits champs récepteurs et des variations lumineuses provenant de petites régions de l'espace suffisent à modifier leur activité. Elles permettent donc le traitement de fréquences spatiales relativement hautes. Par ailleurs, leurs axones ont un petit diamètre et ne sont pas myélinisés. Elles transmettent donc l'information plus lentement que les cellules M, de manière tonique (i.e. elles répondent à une stimulation soutenue). Plus précisément, des enregistrements neurophysiologiques chez le macaque suggèrent que l'information issue des cellules P atteindrait le cortex visuel primaire environ 20 ms après l'information issue des cellules M (Nowak & Bullier 1997; Nowak, Munk, Girard, & Bullier 1995). Les cellules P sont moins sensibles au contraste de luminance mais permettent de traiter les variations chromatiques. À noter que la répartition des cellules sur la rétine n'est pas homogène, de telle sorte que le centre de la rétine (fovéa) est principalement composé de cônes dont le signal est transmis principalement à des cellules P, tandis que la zone périphérique de la rétine contient essentiellement des bâtonnets dont le signal est transmis principalement à des cellules M (Curcio & Allen 1990; Wässle, Grünert, Röhrenbeck, & Boycott 1990). Nous reviendrons plus loin dans le manuscrit sur cette propriété de la rétine fondamentale à la problématique de cette thèse.

Les axones des cellules ganglionnaires forment ensuite le nerf optique, qui véhicule l'information jusqu'aux corps genouillés latéraux au sein des thalami. Les caractéristiques de l'information transmise par les cellules ganglionnaires M et P sont préservées dans ces structures puisque les deux types de cellules projettent l'information respectivement vers les couches magnocellulaires et parvocellulaires des corps genouillés latéraux. À la sortie des corps genouillés latéraux, les projections vont être à l'origine de deux voies distinctes menant aux aires corticales visuelles : la voie magnocellulaire et la voie parvocellulaire. C'est le cortex visuel primaire (V1, ou cortex strié) qui reçoit d'abord les projections des corps genouillés latéraux et qui constitue le point de départ d'une série d'aires visuelles organisées de façon plus ou moins hiérarchique. Il est situé dans la partie postérieure des lobes occipitaux et s'étend le long de la scissure calcarine. Au sein de V1, la distinction entre les voies magnocellulaire et parvocellulaire reste préservée. Le travail des cellules de V1 serait de décomposer le signal en tout un tas de primitives visuelles rudimentaires.

En effet, ces cellules se comportent comme des filtres de Gabor orientés, c'est-à-dire qu'elles permettent de coder l'information à différentes orientations et différentes échelles spatiales (De Valois, Albrecht, & Thorell 1982; De Valois & De Valois 1980; Hubel & Wiesel 1959, 1968). Autrement dit, chacune des cellules de V1 code pour une fréquence spatiale et une orientation donnée.

Les propriétés neurophysiologiques des cellules de la rétine au cortex, et leur organisation, indiquent trois choses importantes : **(i)** l'information issue des basses et des hautes fréquences spatiales est véhiculée jusqu'au cortex *via* deux voies distinctes (les voies magnocellulaire et parvocellulaire, respectivement), **(ii)** les propriétés de réponse des cellules formant ces deux voies sont telles que l'information se propage plus rapidement dans la voie magnocellulaire que dans la voie parvocellulaire et **(iii)** les cellules du cortex visuel primaire répondent sélectivement à différentes orientations et fréquences spatiales. On a donc un système visuel qui est spécifiquement équipé pour traiter différentes fréquences spatiales en parallèle, et les larges blobs de luminance décrivant la structure de la scène, portés par les BFs, sont théoriquement disponibles plus rapidement que les contours fins et les détails, portés par les HFs. Comme nous l'avons déjà évoqué, les propriétés temporelles du traitement des fréquences spatiales ont donc amené à formuler l'hypothèse d'une séquence de traitement *coarse-to-fine* de l'information visuelle.

2.2. Le modèle fréquentiel

Le modèle fréquentiel a été proposé par Christian Marendaz (Marendaz et al. 2003) afin de rendre compte de la rapidité et de l'efficacité de la reconnaissance visuelle. Ce modèle est directement dérivé du principe de traitement *coarse-to-fine* (initialement proposé par Schyns et Oliva (1994)), mais il permet d'insérer ce principe au sein d'un modèle qui considère les conséquences d'une séquence de traitement *coarse-to-fine* pour la reconnaissance visuelle. Plus exactement, le modèle fréquentiel postule **(i)** que les relations spatiales entre des blobs de luminance (i.e. les BFs) domineraient le traitement précoce de la scène et permettraient une première catégorisation et **(ii)** que cette première catégorisation serait ensuite affinée, validée ou infirmée par l'accumulation progressive d'information en HFs. Les premiers traitements visuels dans ce modèle correspondent ainsi à une reconnaissance visuelle ultra-rapide et à une perception du *gist* de la scène. Le modèle fréquentiel est corroboré par de nombreuses données empiriques, qui sont en partie exposées ci-dessous.

2.2.1. Le phénomène de précédenace globale et les formes hiérarchiques

Avant que l'on ne s'intéresse aux fréquences spatiales comme caractéristiques du signal visuel fondamentales à la reconnaissance visuelle, l'idée d'un avantage du traitement de l'information disponible au niveau global par rapport au niveau local était déjà une idée importante en psychologie. Elle est par exemple au cœur de la théorie de la Gestalt (ou psychologie de la forme), qui a marqué la psychologie expérimentale européenne au

20^{ème} siècle. Dans une formule notoire, les théoriciens de la Gestalt postulaient que le tout est supérieur à la somme des parties, et ainsi que la perception d'un objet dans sa globalité prévaut sur celle des parties qui le composent. Le système organiserait les éléments visuels selon un certain nombre de lois afin de former un tout cohérent. Par exemple, les composantes visuelles seraient regroupées selon la proximité ou la similarité entre les éléments. Ce n'est d'ailleurs pas sans rappeler les percepts d'ensemble créés par le phénomène de *summary statistics* (le percept d'un mur ou bien d'un buisson semble plus important que celui des briques ou des feuilles qui les composent, respectivement).

La primauté de la perception globale sur la perception locale a été mise en évidence expérimentalement à l'aide du paradigme des formes hiérarchiques, qui permet de montrer que la structure globale d'un stimulus est identifiée plus rapidement que les éléments qui la composent. Les stimuli hiérarchiques sont constitués d'une forme globale composée de plusieurs formes locales (e.g., Figure 1.7-A). À titre d'exemple, Navon (1977) utilisait des lettres globales (e.g., un grand H) composées de plusieurs petites lettres (e.g., des petits s) et observait que l'identification de la lettre globale était plus rapide que celle de la lettre locale. On observe également grâce à ce type de stimuli que lorsque la tâche exige un traitement local, celui-ci est gêné par l'information globale traitée rapidement et automatiquement. On parle alors d'un effet d'interférence globale. Ce phénomène n'est pas observé dans le cas contraire : lorsque la tâche exige un traitement global, celui-ci n'est pas gêné par la présence des éléments locaux (Kimchi 1992; Navon 1977). Par ailleurs, une autre expérience montrait que lorsque les participants devaient identifier des lettres en présentation auditive, tandis qu'ils percevaient visuellement les stimuli hiérarchiques, seule la lettre globale interférait sur la reconnaissance du stimulus auditif (Navon 1977). Ces résultats sont toujours considérés comme une démonstration importante d'une précedence du traitement global. Cependant, comme Kimchi (1992) l'a justement souligné, l'hypothèse de précedence globale concerne le développement du percept dans le temps : on percevrait d'abord la structure globale de notre environnement avant que des traitements locaux puissent être réalisés. En revanche, l'hypothèse de précedence globale n'indique rien sur la disponibilité des informations locales et globales dans le percept final. Autrement dit, elle ne permet pas de dire ce qui est le plus saillant (structure globale ou éléments locaux) dans le percept final, ou comment celui-ci est interprété.

L'effet d'interférence globale est important à considérer car il indique qu'il n'y a pas d'indépendance entre les traitements réalisés au niveau global et ceux réalisés au niveau local. Mais à quel niveau surviennent ces interactions ? Pour Nicolas Poirel et ses collaborateurs (Poirel et al. 2010; Poirel, Mellet, Houdé, & Pineau 2008; Poirel, Pineau, & Mellet 2006), l'effet d'interférence globale serait étroitement lié à l'analyse sémantique du stimulus visuel. Ces auteurs ont réalisé une série d'expériences dans lesquelles ils utilisaient des formes hiérarchiques dont la significativité était systématiquement manipulée. Dans l'une d'entre elles par exemple, le niveau global et le niveau local de chaque

stimulus pouvait être constitué d'objets, ou bien de formes sans signification ("non-objets", [Poirel et al. 2008](#)). Lors d'un essai, une forme hiérarchique était présentée. Les participants devaient porter leur attention soit au niveau global soit au niveau local, et indiquer si une cible y était présente ou non. La cible pouvait être un objet ou un non-objet, et de la même manière, la forme au niveau non pertinent pouvait être un objet ou un non-objet (Figure 1.7-B). Les résultats montraient d'abord que les participants étaient plus rapides pour réaliser la tâche au niveau global qu'au niveau local, témoignant d'un effet de précedence globale. Ensuite, lorsque les participants réalisaient la tâche au niveau global, les temps de réponse étaient similaires entre la condition où les formes locales étaient des objets et celle où les formes locales étaient des non-objets. Par contre, lorsqu'ils réalisaient la tâche au niveau local, le traitement des formes locales était gêné par la forme globale, et cet effet d'interférence globale était plus grand lorsque la forme globale représentait un objet que lorsqu'il s'agissait d'un non-objet. Ainsi, un stimulus porteur de sens au niveau global capturerait l'attention avec une telle force qu'il gênerait d'autant plus le traitement de l'information locale.

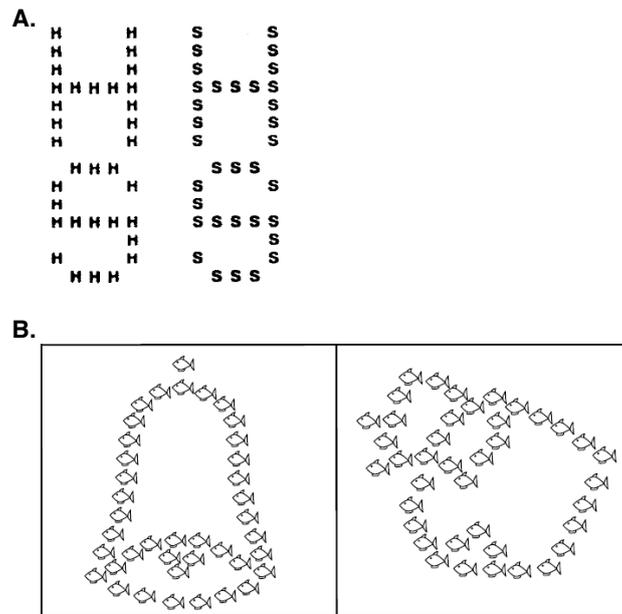


Figure 1.7: **A.** Stimuli hiérarchiques. Adapté de [Kimchi \(1992\)](#). **B.** Exemple de stimuli utilisés dans l'étude de [Poirel et al. \(2008\)](#). Ici, la cible est un poisson et doit être identifiée au niveau local. Dans cette condition, les temps de réponse sont plus longs lorsque la forme globale représente un objet identifiable (à gauche) qu'un non-objet (à droite).

Dans les années 1980 et 1990, certains chercheurs se sont demandé si l'effet de précedence globale ne résultait pas du traitement rapide de l'information globale contenue dans les BFs ([Badcock, Whitworth, Badcock, & Lovegrove 1990](#); [Lamb & Yund 1993](#)). Cette hypothèse découle en fait de travaux psychophysiques qui ont démontré une correspondance entre les canaux fréquentiels et les niveaux de traitement global/local de l'information visuelle. Cette correspondance a par exemple été mise en évidence par

Shulman, Sullivan, Gish, et Sakoda (1986) en exploitant le phénomène d'adaptation visuelle. L'adaptation se définit comme un changement temporaire de la sensibilité ou de la perception subjective après exposition à un nouveau stimulus (Thompson & Burr 2009; Webster 2011), qui révèle la capacité du système visuel à s'ajuster aux changements dans l'environnement. En utilisant ce principe, il est possible de provoquer une adaptation sélective à une fréquence spatiale donnée. Le principe est de présenter de façon prolongée un réseau sinusoïdal (ou *grating*) à un observateur. Un réseau sinusoïdal est un stimulus constitué d'une seule sinusoïde et qui ressemble donc à une alternance progressive de bandes sombres et de bandes claires dont la luminance respective définit le contraste du réseau (Figure 1.8). Après avoir été adapté à un réseau d'une fréquence spatiale donnée, la détection d'un réseau de fréquence spatiale identique requiert un plus haut niveau de contraste (alors que le seuil de détection de réseaux à d'autres fréquences spatiales n'est pas affecté; Blakemore & Campbell 1969). Shulman et al. (1986) ont mesuré les performances de participants sur une tâche de stimuli hiérarchiques avant et après les avoir adapté à des réseaux sinusoïdaux en BFs ou HF. L'adaptation aux BFs diminuait les performances au niveau global tandis que l'adaptation aux HF diminuait les performances au niveau local. Ce résultat indique que l'information locale serait portée par les HF et l'information globale par les BFs.

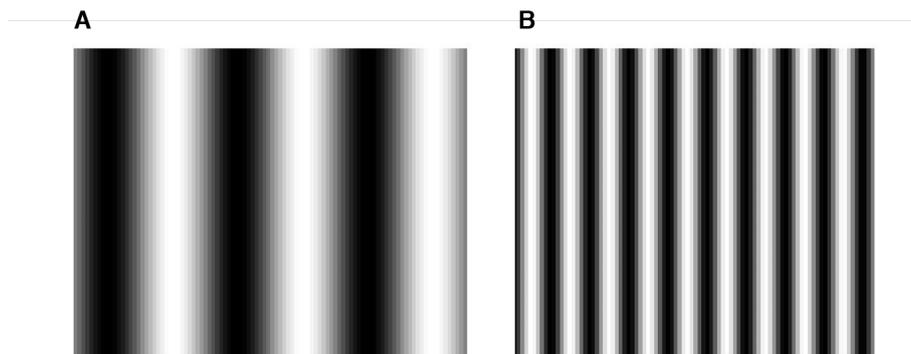


Figure 1.8: Réseaux sinusoïdaux. **A.** Basses fréquences spatiales (3 cycles par image) et **B.** Hautes fréquences spatiales (10 cycles par image)

Confortés par ce résultat, Badcock et al. (1990) ont directement testé la contribution des fréquences spatiales dans l'effet de précedence globale. Ils présentaient à des participants un stimulus hiérarchique (une grande lettre composée de petites lettres, comme dans l'expérience princeps de Navon) soit avec toutes ses composantes fréquentielles, soit après avoir appliqué un filtre-passe haut qui préservait les HF mais supprimait le signal en BFs. Lorsque les stimuli n'étaient pas filtrés, les auteurs observaient de nouveau un effet de précedence globale. La lettre globale était identifiée plus rapidement que les petites lettres qui la composaient. Mais le résultat substantiel de cette étude est que lorsque les BFs étaient retirées du stimulus (stimuli avec le filtre passe-haut), la précedence globale n'était plus observée. Ce résultat suggère que l'effet de précedence

globale reposerait sur le traitement rapide des BFs. Pour conclure sur l'ensemble des résultats obtenus avec les formes hiérarchiques, on peut dire que le modèle fréquentiel peut constituer un bon moyen de rendre compte de l'effet de précedence globale.

2.2.2. Précedence du traitement des BFs dans la perception de réseaux sinusoïdaux et de scènes

La précedence du traitement des BFs par rapport à celui des HFes a été mise en évidence de nombreuses fois en manipulant explicitement les fréquences spatiales. L'exemple le plus simple est celui de la détection de réseaux sinusoïdaux. En mesurant le temps nécessaire pour détecter des réseaux sinusoïdaux, des études psychophysiques ont montré que le temps de détection des réseaux augmentent avec l'augmentation de la fréquence spatiale (Breitmeyer 1975; Musselwhite & Jeffreys 1985; Parker & Dutch 1987). Ces résultats sont corroborés par des études en potentiels évoqués utilisant des réseaux sinusoïdaux de différentes fréquences spatiales. En particulier, Parker et Salzen (1977) ont observé une augmentation progressive de la latence de potentiels évoqués visuels avec l'augmentation de la fréquence spatiale (voir aussi Parker, Salzen, & Lishman 1982).

L'étude princeps de Schyns et Oliva (1994) que nous avons décrite plus haut fait bien sûr écho à ces résultats. Pour rappel, le paradigme expérimental utilisé dans cette étude consistait à présenter d'abord une scène hybride - comprenant les BFs d'une scène superposées aux HFes d'une autre scène - puis une scène non filtrée, et les participants devaient indiquer si la scène non filtrée était présente ou non dans l'hybride. Les résultats indiquaient que l'interprétation qui est faite de la scène hybride est plus souvent basée sur les BFs lorsque l'hybride était présenté pendant 30 ms tandis qu'elle est plus souvent basée sur les HFes lorsque l'hybride est présenté pendant 150 ms. Ainsi, la précedence de traitement des BFs comparativement aux HFes n'est pas un phénomène restreint à des stimuli simples comme des formes hiérarchiques ou des réseaux sinusoïdaux. Les photographies de scènes naturelles, composées d'une multitude de sinusoïdes, enclenchent les mêmes propriétés temporelles de traitement. La scène serait alors analysée selon une séquence *coarse-to-fine*. Selon Schyns et Oliva (1994), ce schéma d'analyse pourrait aussi indiquer que la reconnaissance de la scène serait possible très rapidement, alors que la reconnaissance des objets dans la scène serait secondaire et dominerait pendant les stades tardifs du traitement.

Dans une seconde expérience, Schyns et Oliva (1994) ont testé directement l'hypothèse d'une stratégie d'analyse *coarse-to-fine* en faisant un usage ingénieux des stimuli hybrides. Cette fois, chaque stimulus consistait en une séquence de deux images hybrides présentées successivement et rapidement (45 ms chacune). Chaque séquence était construite de telle sorte que les BFs de la première image hybride et les HFes de la seconde image hybride appartenaient à la même scène, et *vice versa* (comme si l'on présentait les deux hybrides de la Figure 1.2 successivement). Les participants n'étaient pas informés de cette manipulation, et ils avaient pour tâche de catégoriser chaque stimulus

selon quatre catégories de niveau de base (salon, chambre, ville et autoroute). Leur réponse permettait de déterminer s'ils avaient catégorisé l'hybride en suivant une séquence *coarse-to-fine* ou *fine-to-coarse*. La logique était la suivante : s'ils donnaient pour réponse la catégorie de la scène correspondant aux BFs de la première image hybride et aux HFs de la seconde (par exemple, *autoroute* dans la Figure 1.2), cela signifiait qu'ils avaient traité le stimulus selon une séquence *coarse-to-fine*. À l'inverse, s'ils donnaient pour réponse la catégorie de la scène correspondant aux HFs de la première image hybride et aux BFs de la seconde (*ville* dans la Figure 1.2), cela signifiait qu'ils avaient traité le stimulus selon une séquence *fine-to-coarse*. Les résultats ont montré que 67% des stimuli étaient catégorisés selon une séquence de traitement *coarse-to-fine*, contre seulement 29% selon une séquence *fine-to-coarse* (les 4% restant correspondaient à des erreurs de catégorisation).

Avec le même objectif de tester l'hypothèse de l'utilisation préférentielle d'une séquence de traitement *coarse-to-fine*, des études plus récentes réalisées dans l'équipe de Carole Peyrin au Laboratoire de Psychologie et NeuroCognition (LPNC) ont utilisé une méthodologie différente qui consiste à manipuler l'ordre de présentation des basses et hautes fréquences spatiales de scènes (Kauffmann, Chauvin, Guyader, & Peyrin 2015a; Kauffmann, Chauvin, Pichat, & Peyrin 2015b; Musel, Chauvin, Guyader, Chokron, & Peyrin 2012; Musel et al. 2014). Le paradigme expérimental consiste à présenter des séquences composées de six images filtrées d'une même scène (filtrage passe-bande) et assemblées en allant des BFs aux HFs ou des HFs aux BFs, permettant ainsi d'imiter et d'imposer au système visuel une analyse *coarse-to-fine* ou une analyse *fine-to-coarse*, respectivement. Les résultats de ces études ont montré que les temps de réponse pour catégoriser ces stimuli étaient plus rapides lorsque les scènes étaient assemblées suivant une séquence *coarse-to-fine* que *fine-to-coarse*, suggérant que le système visuel tire profit d'avoir accès en premier lieu aux BFs d'une scène. À noter qu'il s'agissait cette fois-ci d'une catégorisation au niveau sur-ordonné puisque les participants devaient catégoriser les scènes comme étant des scènes d'intérieur ou d'extérieur.

2.2.3. Limites du modèle fréquentiel : Un principe par défaut

Il est nécessaire d'émettre des réserves sur la domination des BFs lors de la catégorisation de stimuli visuels. Les résultats de plusieurs études comportementales ont en effet souligné la flexibilité du traitement visuel *coarse-to-fine* (Bonnar, Gosselin, & Schyns 2002; Oliva & Schyns 1997; Özgen, Sowden, Schyns, & Daoutis 2005; Schyns, Bonnar, & Gosselin 2002; Schyns & Oliva 1999). Par exemple, dans l'étude de Schyns et Oliva (1994), on remarque que bien que la stratégie *coarse-to-fine* était privilégiée dans la majorité des cas (67%), environ 30% des catégorisations reflétaient tout de même une stratégie *fine-to-coarse*. Une autre étude des mêmes auteurs (Schyns & Oliva 1999) suggère que l'information recherchée pour répondre à des tâches de catégorisation de visages pouvait contraindre le système visuel à utiliser les HFs plutôt que les BFs d'une image hybride. Plus précisément, les résultats de cette étude montrent que les HFs de

l'hybride seraient préférentiellement utilisées pour décider si un visage est expressif ou non, alors que les BFs seraient préférentiellement utilisées pour catégoriser l'expression en question (i.e. heureux, en colère, neutre). Similairement, [Rotshtein, Schofield, Funes, et Humphreys \(2010\)](#) ont montré que la bande de fréquences spatiales utile à la catégorisation d'un des deux composants d'une image hybride dépend directement du type de catégorisation. Les BFs seraient favorisées pour catégoriser un visage *vs* une fleur, mais les HFes seraient favorisées pour catégoriser un visage *vs* une maison. L'analyse *coarse-to-fine* pourrait donc correspondre à une analyse "par défaut" efficace en situation de perception passive, mais ce principe serait flexible afin de permettre un traitement plus rapide des détails portés par les HFes lorsque ceux-ci sont utiles pour répondre à certaines tâches de catégorisation. [Schyns et Oliva \(1997\)](#) avaient alors proposé que des contraintes de haut niveau (i.e. liées à la tâche) pouvaient avoir un rôle direct sur la sélection de l'information en fréquences spatiales jugée la plus diagnostique pour réaliser la tâche en question.

Pour résumer, le modèle fréquentiel propose que grâce à un traitement parallèle des BFs et des HFes, le système visuel peut rapidement capturer le *gist* de la scène afin de faire une première catégorisation qui serait ensuite affinée par l'arrivée plus tardive des HFes. Il y aurait donc un principe d'intégration graduelle et ascendante de l'information visuelle des BFs aux HFes (*coarse-to-fine*), qui pourrait néanmoins être flexible dans le cas où la tâche requiert un traitement plus précoce des HFes. Mais dans la lignée du modèle fréquentiel et sous l'influence grandissante des théories du codage prédictif, il a été proposé que le traitement des BFs permettrait en fait de guider activement et de faciliter l'analyse plus tardive des HFes *via* des mécanismes prédictifs fonctionnant de manière descendante (*top-down*). La rapidité et l'efficacité de la reconnaissance visuelle serait alors due à l'intégration d'un signal descendant véhiculant les résultats de la première analyse du *gist* de la scène, grâce au traitement des BFs, à un signal ascendant correspondant au traitement des HFes. C'est la proposition d'un certain nombre de chercheurs ([Bar 2003, 2007](#); [Kauffmann, Ramanoël, & Peyrin 2014](#); [Peyrin et al. 2010](#); [Trapp & Bar 2015](#)), qui a notamment été formalisée par Moshe Bar dans un modèle neurofonctionnel proactif de la reconnaissance visuelle ([Bar 2003, 2007](#)). L'hypothèse d'influences prédictives descendantes dans la reconnaissance visuelle est appuyée par un certain nombre d'arguments issus de la neurophysiologie et de la neuroimagerie. Nous les passerons en revue avant de présenter le modèle proactif.

2.3. Influences descendantes et mécanismes prédictifs du traitement visuel

2.3.1. Le cortex visuel : Une architecture hiérarchique mais des connexions bidirectionnelles

Nous l'avons évoqué plus haut, le cortex visuel primaire (ou strié) est considéré comme étant le point de départ d'une série d'aires visuelles organisées de façon hiérarchique. Les cellules du cortex visuel primaire projettent de manière ascendante vers le cortex extra-strié, à partir duquel on observe une séparation des traitements en deux voies aux propriétés fonctionnelles différentes, la voie ventrale et la voie dorsale (Goodale & Milner 1992; Ungerleider & Haxby 1994). La voie dorsale véhicule l'information visuelle vers le cortex pariétal postérieur, notamment *via* l'aire MT (V5 chez le primate non-humain), une aire visuelle responsable de la perception du mouvement. La majorité des fibres qui se projettent dans la voie dorsale sont des fibres magnocellulaires. La latence de réponse des aires dorsales est donc courte, et ces aires présentent une bonne sensibilité au mouvement (notamment dans l'aire MT) et aux informations achromatiques en BF. Les lésions dans la voie dorsale ont ainsi tendance à provoquer des déficits sur des tâches impliquant le traitement de relations spatiales et du mouvement. La voie ventrale véhicule quant à elle l'information du cortex occipital ventral jusque dans le cortex inférotemporal. Elle reçoit à la fois des afférences magnocellulaires et parvocellulaires, avec une proportion plus importante d'afférences parvocellulaires. Les aires de la voie ventrale sont sensibles aux informations de chrominance (notamment dans l'aire V4), aux formes, et à un spectre large de fréquences spatiales, des BF aux HF. Les lésions dans la voie ventrale provoquent ainsi des déficits pour le traitement des couleurs (achromatopsie), la reconnaissance des objets (agnosie) ou spécifiquement des visages (prosopagnosie).

La distinction voie dorsale/voie ventrale a tout d'abord été expérimentalement mise en évidence à travers l'observation de comportements induits par des lésions corticales chez le singe (e.g., Mishkin, Ungerleider, & Macko 1983) et chez l'homme (e.g., Newcombe, Ratcliff, & Damasio 1987; Ungerleider & Haxby 1994). À titre d'exemple, Newcombe et al. (1987) ont observé une double-dissociation entre la reconnaissance visuelle et les performances visuo-spatiales. Ils reportent le cas d'un patient avec une lésion du cortex occipito-temporal qui présentait un déficit dans une tâche de reconnaissance de visages (*Mooney faces*) mais avait des performances normales dans une tâche de résolution de labyrinthe, et le cas d'un second patient, avec une lésion du cortex occipito-pariétal, qui présentait le pattern inverse. Pour résumer, on considère classiquement que la voie ventrale est impliquée dans le traitement des couleurs et dans la reconnaissance visuelle alors que la voie dorsale serait plutôt impliquée dans les traitements spatiaux (Goodale & Milner 1992; Ungerleider & Haxby 1994).

Le long de la voie ventrale, on constate une complexité grandissante des traitements du cortex visuel primaire au cortex inférotemporal. Dans le cortex visuel primaire, les

cellules ont des champs récepteurs restreints et répondent, comme nous l'avons vu, à des fréquences spatiales et des orientations données en fonction de leur position dans le champ visuel. La taille des champs récepteurs augmente ensuite avec la hiérarchie corticale. Autrement dit, les aires de plus haut niveau dans la hiérarchie peuvent intégrer l'information provenant de régions plus distantes du champ visuel que les aires de plus bas niveau. Dans les aires intermédiaires, les cellules répondent à des caractéristiques visuelles de niveau intermédiaire, comme les contours (Pasupathy & Connor 1999). Plus loin dans la hiérarchie corticale, les champs récepteurs du cortex inférotemporal sont larges (possiblement $10^\circ \times 10^\circ$ d'angle visuel) et intègrent les contours en des formes complexes (Brincat & Connor 2004). Une illustration de cette complexité progressive se trouve au sein d'une région latérale du cortex occipito-temporal, le LOC (pour *Lateral Occipital Complex*). Le LOC est une région fonctionnellement définie comme répondant plus fortement aux objets qu'à d'autres stimuli, comme des textures, des visages ou des scènes, et qui s'étend sur la partie latérale du gyrus fusiforme et du gyrus temporal inférieur (Grill-Spector, Kourtzi, & Kanwisher 2001; Grill-Spector, Kushnir, Hendler, et al. 1998). Au sein du LOC, les sous-régions les plus antérieures représentent des percepts plus invariants à la position, à l'orientation et aux caractéristiques de bas niveau que les sous-régions les plus postérieures (Grill-Spector et al. 1999; Grill-Spector, Kushnir, Edelman, Itzhak, & Malach 1998). Les traitements opérés dans la voie ventrale procéderaient alors de façon hiérarchique en intégrant peu à peu l'information, de caractéristiques très locales à un percept global et invariant.

Une conception hiérarchique des traitements visuels (du plus simple au plus complexe), bien que soutenue par les propriétés fonctionnelles du cortex visuel que nous venons de décrire, peut difficilement rendre compte de l'effet de précedence globale ou d'une séquence de traitement *coarse-to-fine*, c'est-à-dire d'une dynamique de traitement visuel qui, au contraire, intégrerait l'information d'un niveau global à un niveau local. Ces vues semblent même diamétralement opposées. Cependant, elles sont réconciliables si on considère deux autres propriétés du cortex visuel. Tout d'abord, il existe des connexions descendantes entre les aires visuelles dont le nombre serait même plus important que le nombre de connexions ascendantes (Bullier 2001). Ensuite, les travaux en neurophysiologie chez le macaque montrent que les latences de réponse des cellules d'aires de haut niveau dans la hiérarchie sont souvent similaires, et parfois plus courtes, que celles d'aires de plus bas niveau (Bullier 2001). On remarque que les aires qui ont les latences de réponse les plus courtes sont situées sur la voie dorsale (ou dans le prolongement de cette voie dans le cortex frontal) et reçoivent alors principalement des projections magnocellulaires. Par exemple, chez le macaque, les latences de réponse des cellules du *Frontal Eye Field*, une aire du cortex frontal située dans le prolongement de la voie dorsale et impliquée dans les traitements oculomoteurs, ou encore celles de l'aire MT/V5 qui code l'information de mouvement, seraient très comparables aux latences de réponse des cellules de V1 (Figure 1.9).

Bullier (2001) a alors proposé un modèle intégratif du traitement visuel qui permet

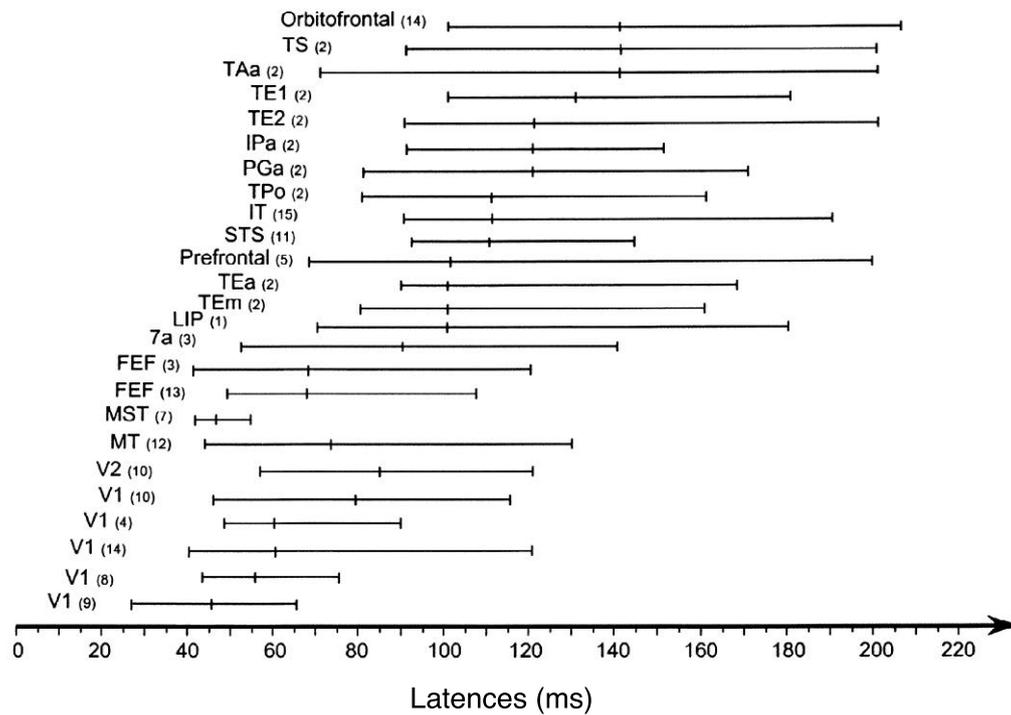


Figure 1.9: Latences de réponse visuelle de neurones dans différentes aires corticales (Bullier 2001). Pour chaque aire, la barre verticale centrale représente la latence médiane et les barres verticales de chaque côté représentent les 10^e et 90^e centiles. Les aires sont organisées plus ou moins hiérarchiquement du bas vers le haut (de V1 à des aires frontales comme le cortex orbitofrontal). On remarque que les latences de réponse d'aires hautes dans la hiérarchie ne sont pas nécessairement plus grandes que celles d'aires de plus bas niveau. Par exemple, les latences du cortex inférotemporal (IT), des *Frontal Eye Fields* (FEF) ou du cortex préfrontal recouvrent partiellement celles de V1.

de tenir compte de ces particularités. Dans ce modèle, la propagation de l'information visuelle ne serait pas purement ascendante mais procéderait par des vagues d'activation successives. Une première vague d'activation serait déclenchée par l'arrivée d'un signal magnocellulaire qui se propagerait à travers la voie dorsale afin d'activer des aires de haut niveau. Une seconde vague d'activation correspondrait à la rétro-injection sur des aires de plus bas niveau du résultat des traitements de la première vague. Une troisième vague d'activation serait déclenchée par l'arrivée d'un signal parvocellulaire sur les aires de bas niveau. Ces aires de bas niveau auraient alors un rôle d'intégration entre l'information parvocellulaire ascendante et les résultats de l'analyse de l'information magnocellulaire descendante. Bullier (2001) ne donne pas de précision sur le type d'informations véhiculées dans ce modèle. Cependant, du fait des propriétés temporelles de réponse des cellules M (qui répondent aux BF) et P (qui répondent aux HF), le modèle pourrait très bien prédire une influence active du résultat du traitement des BF sur le traitement des HF. Plus précisément, l'information grossière et globale portée par les BF pourrait être rétro-injectée pour contraindre le traitement de l'information locale portée par les

HFs.

À noter que Bullier (2001) insiste particulièrement sur le rôle d'intégration de V1 et V2. Néanmoins, il semble raisonnable de penser que d'autres aires plus hautes hiérarchiquement puissent aussi jouer ce rôle d'intégration. Les aires du cortex inférotemporal, comme le LOC, pourraient par exemple tout aussi bien intégrer l'information. Cette conception plus générale d'influences descendantes (ou *feedback*) des aires hiérarchiquement supérieures sur les aires hiérarchiquement inférieures est au cœur des théories prédictives du traitement visuel.

2.3.2. Codage prédictif et modèles neurocognitifs prédictifs de la reconnaissance visuelle

L'influence de la théorie du codage prédictif sur les champs de la psychologie, de la psychiatrie et des neurosciences est croissante depuis une vingtaine d'années. Globalement, le postulat de cette théorie est le suivant. Le cerveau construit peu à peu des modèles internes du monde extérieur qui sont utilisés pour prédire les entrées sensorielles. Lorsqu'une afférence sensorielle entre dans le système, tout ce qui ne peut pas être prédit sur la base des modèles internes est transmis aux aires corticales supérieures sous la forme d'un signal de surprise, ou d'erreur de prédiction. L'intérêt d'un tel système est qu'il permet d'interpréter des entrées sensorielles bruitées. Par exemple, si mon modèle interne du monde contient l'information que mes chaussures sont sur le tapis de l'entrée, je vais être capable de reconnaître instantanément mes chaussures dans l'entrée, même si celles-ci sont partiellement masquées par le chat qui passe devant à ce moment-là. Des mécanismes de codage prédictif permettraient ainsi d'expliquer comment l'information sensorielle manquante est complétée, comme dans le phénomène des contours illusoire observé par exemple avec le triangle de Kanizsa (Figure 1.10 ; voir par exemple Kok, Bains, van Mourik, Norris, & de Lange 2016). Parmi les importants arguments empiriques à l'appui du principe de codage prédictif, on distingue notamment les phénomènes de *mismatch negativity*, une réponse électrophysiologique à la surprise. Par exemple, lorsqu'un sujet entend une suite de sons similaires qui se répètent de manière régulière, l'apparition d'un son déviant, donc inattendu, déclenche un changement large dans les potentiels évoqués auditifs (Näätänen, Paavilainen, Rinne, & Alho 2007). Il a aussi été démontré que plus la probabilité d'apparition du son déviant augmente et plus l'amplitude du *mismatch negativity* diminue (Näätänen & Picton 1987).

Dans le domaine visuel, c'est sans doute l'article de Rao et Ballard (1999) qui constitue la pierre angulaire des théories du codage prédictif. Ces auteurs ont proposé un modèle du traitement visuel analogue à celui de Bullier (2001), mais qui introduit en outre la notion de prédiction (et d'erreur de prédiction). Dans ce modèle, les aires corticales supérieures essaient de prédire activement l'activité des aires corticales inférieures *via* les connexions descendantes. Les connexions ascendantes serviraient quant à elles à véhiculer l'erreur de prédiction (i.e. tout ce qui n'a pas pu être prédit), tout ce qui était prédictible n'étant alors pas transmis. L'erreur de prédiction pourrait ainsi permettre

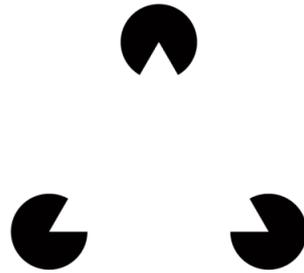


Figure 1.10: Triangle de Kanizsa, reflétant le phénomène de contours illusoires dans lequel la perception d'un contour est possible sans changement de luminance. Lors de la perception d'une telle figure, les neurones de V1 dont les champs récepteurs se situent sur le contour illusoire s'activent (T. S. Lee & Nguyen 2001). Ce phénomène pourrait s'expliquer par des mécanismes de codage prédictif impliquant l'arrivée d'un signal descendant sur ces cellules (Kok et al. 2016).

de mettre à jour les modèles internes de l'environnement codés dans les aires corticales supérieures. Le modèle de Rao et Ballard a déclenché plusieurs propositions théoriques parmi lesquelles on distingue celle, très influente, de Karl Friston (2009; 2010; Friston & Stephan 2007) sur la minimisation de l'énergie libre, ou encore indirectement celle du modèle proactif de la reconnaissance visuelle de Moshe Bar (2003; 2007; Bar et al. 2006), qui n'est pas un modèle de codage prédictif en soi, mais qui a sans doute été influencé par ce cadre de pensée.

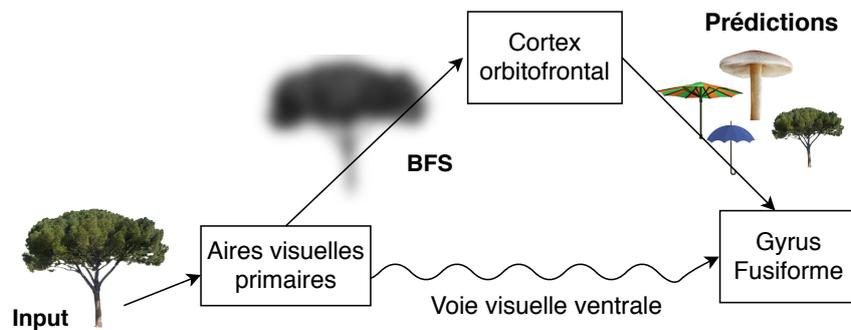


Figure 1.11: Modèle neurocognitif proactif de Moshe Bar. L'information grossière contenue dans les BF de l'objet est projetée rapidement vers le cortex orbitofrontal, permettant d'activer des prédictions. Parallèlement, l'information visuelle se propage le long de la voie ventrale jusqu'au gyrus fusiforme dans le cortex inférotemporal. Les prédictions sont ensuite projetées sur le gyrus fusiforme et intégrées à l'analyse ascendante, afin de faciliter la reconnaissance. Adapté de Bar (2003) et Bar et al. (2006).

Le modèle proactif de Bar postule que la reconnaissance d'un objet fait intervenir trois mécanismes successifs (Figure 1.11) : (i) les BF sont d'abord projetées rapidement vers le cortex préfrontal *via* les fibres magnocellulaires le long de la voie dorsale. Cette

représentation grossière permet d'activer des prédictions à propos de l'identité de l'objet. Par exemple, la forme grossière dans la Figure 1.11 pourrait être un champignon, une lampe ou encore un parapluie, mais difficilement une voiture ou un ordinateur. **(ii)** Ces prédictions sont ensuite transmises vers les aires du cortex occipito-temporal, comme le gyrus fusiforme, afin de limiter les candidats à la reconnaissance. **(iii)** L'intégration des prédictions et des traitements ascendants réalisés le long de la voie ventrale résulte en la reconnaissance finale de l'objet. L'information en BFs serait ainsi suffisante pour générer des prédictions et permettre de limiter le nombre de candidats à la reconnaissance dans le cortex occipito-temporal.

Une étude de Bar et al. (2006) qui combine mesures d'IRM fonctionnelle (IRMf) et de magnétoencéphalographie (MEG) a permis d'appuyer le modèle proactif. Dans la première expérience de cette étude, les participants devaient reconnaître des dessins d'objets, présentés très rapidement (26 ms) et suivis d'un masque. Les objets étaient difficilement identifiables la première fois qu'ils étaient perçus, mais la reconnaissance devenait possible avec la répétition des stimuli. Les auteurs ont ensuite comparé l'activité cérébrale induite par les essais pour lesquels les objets avaient été reconnus (i.e. pour lesquelles la réponse comportementale de catégorisation était bonne) à celle induite par les essais pour lesquels ils n'avaient pas été reconnus. Ils ont observé une activation dans le cortex orbitofrontal environ 130 ms après l'onset du stimulus, et cette activation orbitofrontale précédait l'activation du cortex occipito-temporal d'environ 50 ms. Une seconde expérience a permis de montrer que l'activation du cortex orbitofrontal était plus importante lors de la catégorisation d'images d'objets filtrées en BFs qu'en HF. Afin de quantifier les interactions fonctionnelles entre le cortex orbitofrontal et les régions visuelles, les auteurs ont réalisé une analyse de synchronie de phase des signaux MEG. Ces analyses ont mis en évidence que pour les objets filtrés en BFs et pour les images d'objets non filtrées, une forte synchronie se mettait en place entre les régions visuelles occipitales et le cortex orbitofrontal à partir de 80 ms après la présentation du stimulus. Dans un second temps, à partir de 130 ms après la présentation du stimulus, c'était le gyrus fusiforme qui entraînait en synchronie avec le cortex orbitofrontal. Par ailleurs, la synchronie entre ces régions pour les images d'objets HF était significativement moins importante (Figure 1.12) que pour les images en BFs. Ces analyses ne permettent pas de déterminer le sens des interactions entre les régions, mais une interprétation dans le cadre du modèle proactif serait que la première synchronie de phase occipital-orbitofrontal correspondrait à la projection ascendante rapide des BFs et que la seconde synchronie de phase orbitofrontal-fusiforme correspondrait cette fois à la projection descendante des prédictions sur le gyrus fusiforme.

Les résultats de Bar et al. (2006) suggèrent que le cortex orbitofrontal est activé lorsque les objets sont reconnus mais pas lorsqu'ils ne sont pas reconnus. D'autres études corroborent ce résultat. Par exemple, des enregistrements intracellulaires chez le singe ont montré que les neurones du cortex orbitofrontal répondent plutôt à la signification du stimulus qu'à ses caractéristiques physiques (Thorpe, Rolls, & Maddison 1983). Dans

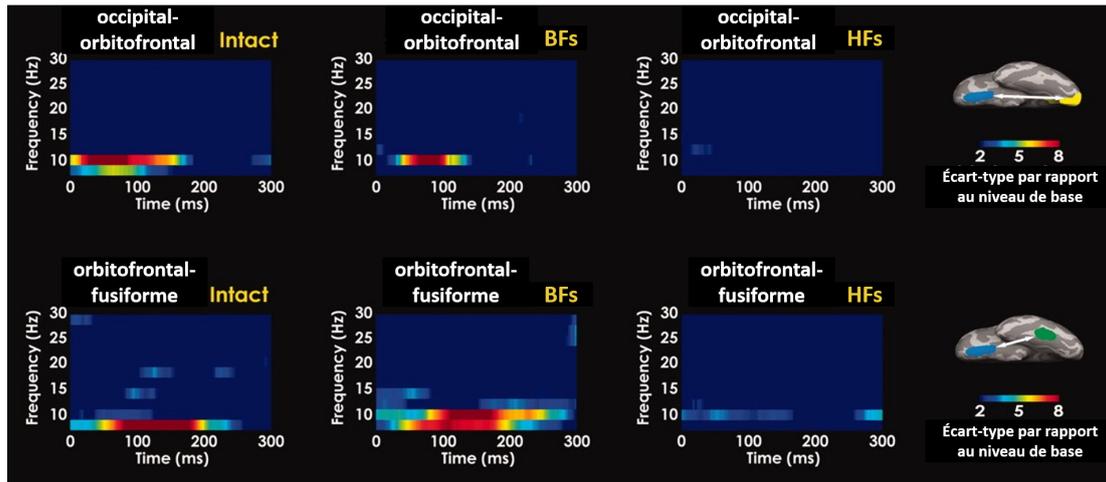


Figure 1.12: Résultats des analyses de synchronie de phase de Bar et al. (2006). Ces résultats suggèrent l’occurrence d’interactions corticales précoces entre les aires visuelles occipitales et le cortex orbitofrontal (en haut), puis entre le cortex orbitofrontal et le gyrus fusiforme (en bas). On constate que ce pattern d’interactions corticales est observé surtout lorsque des BFs sont présentes dans l’image (conditions Intact et BFs). Adapté de Bar et al. (2006).

une étude en IRMf chez l’humain, Chaumon, Kveraga, Barrett, et Bar (2013) ont par ailleurs montré que la perception de patches de Gabor en BFs ou d’objets non identifiables n’activaient pas le cortex orbitofrontal, contrairement à des objets identifiables. L’activité du cortex orbitofrontal et sa connectivité fonctionnelle avec les aires visuelles de la voie ventrale modulait d’ailleurs le degré d’identifiabilité des objets, évalué par un groupe indépendant de participants. Ces résultats suggèrent que des inputs visuels à partir desquels aucune signification ne peut être extraite ne déclenchent pas de prédiction au niveau du cortex orbitofrontal. Par ailleurs, si le cortex orbitofrontal est le site cortical d’un système qui permet la gestion des prédictions sensorielles, il est possible que des prédictions qui ne correspondent pas aux entrées sensorielles causent un surengagement de la région lié à la mise à jour du modèle interne. En effet, plusieurs études montrent que le cortex orbitofrontal serait suractivé en cas d’erreur de prédiction (e.g., Fuster & Bressler 2015; Kauffmann, Bourgin, Guyader, & Peyrin 2015c; Nobre, Coull, Frith, & Mesulam 1999).

Comme le propose Bar (2003; voir aussi Kveraga, Ghuman, & Bar 2007), le cortex orbitofrontal pourrait être rapidement activé grâce aux fibres magnocellulaires à conduction rapide. Cette hypothèse est soutenue par des résultats d’IRMf de la même équipe de recherche (Kveraga, Boshyan, & Bar 2007) qui ont montré que des stimuli biaisés pour activer les cellules M (*M-biased*; i.e. des dessins d’objets achromatiques et faiblement contrastées) activent davantage le cortex orbitofrontal que des stimuli biaisés pour activer les cellules P (*P-biased*; i.e. des dessins d’objets chromatiques et isoluminants vert-rouge). Des analyses supplémentaires ont montré que la magnitude du signal BOLD enregistré dans le cortex orbitofrontal pour les stimuli *M-biased* (mais pas celui

enregistré pour les stimuli *P-biased*) corrélait négativement avec les temps de réponse à la tâche comportementale (moins l'activité du cortex orbitofrontal était forte et plus les participants mettaient de temps pour décider si l'objet en question était plus ou moins grand qu'une boîte à chaussure). Ce résultat suggère qu'il existe une relation entre le système magnocellulaire et l'activité du cortex orbitofrontal. Le signal magnocellulaire pourrait alors être rapidement projeté vers le cortex orbitofrontal afin de déclencher un signal prédictif transmis aux régions visuelles du cortex occipito-temporal (à cet égard, Cavada, Compañy, Tejedor, Cruz-Rizzolo, & Reinoso-Suárez 2000 ont par exemple mis en évidence des connexions anatomiques entre le cortex orbitofrontal et le cortex temporal inférieur chez le macaque). Dans l'étude de Kveraga, Boshyan, et Bar (2007), l'amygdale droite était plus activée pour les stimuli *M-biased* que *P-biased*, contrairement à des régions dorsales (i.e. des régions du cortex pariétal). Les auteurs ont alors proposé que les afférences magnocellulaires pourraient en fait contourner la voie dorsale en étant projetées directement du cortex visuel primaire au cortex orbitofrontal *via* des structures sous-corticales telles que l'amygdale. Cette idée est étayée par des données anatomiques qui montrent que de nombreux axones des neurones de l'amygdale projettent vers le cortex orbitofrontal (Ghashghaei, Hilgetag, & Barbas 2007). Cependant, cette idée reste hypothétique et à notre connaissance, aucune autre donnée n'a depuis permis d'étayer l'hypothèse d'un acheminement partiellement sous-cortical des BF's vers le cortex orbitofrontal.

Si le modèle proactif a été initialement proposé par Moshe Bar pour la reconnaissance d'objets, des études de neuroimagerie ont permis de conforter ce modèle dans le cadre de la perception de scènes naturelles. En particulier, le réseau et la dynamique corticale de l'analyse des fréquences spatiales lors de la catégorisation de scènes ont été plus spécifiquement étudiés par l'équipe de Carole Peyrin au LPNC. Plus précisément, les travaux menés par cette équipe avaient pour objectif d'identifier les régions impliquées dans l'analyse *coarse-to-fine* de scènes, et de déterminer quelles sont les interactions entre ces régions. Dans une première étude en EEG (Peyrin et al. 2010), les participants devaient décider si deux scènes filtrées présentées successivement soit suivant une séquence *coarse-to-fine* (présentation d'une scène en BF's suivie d'une scène en HF's), soit suivant une séquence *fine-to-coarse* (présentation d'une scène en HF's suivie d'une scène en BF's) appartenaient à la même catégorie (plage, ville ou scène d'intérieur). En comparant la source des potentiels évoqués induits par le traitement de ces deux types de séquences, les auteurs ont montré que le traitement de la scène en BF's de la séquence *coarse-to-fine* s'accompagnait d'une augmentation de l'activité du cortex frontal inférieur (en comparaison au traitement de la scène en HF's de la séquence *fine-to-coarse*) entre 150 et 194 ms après l'apparition de la scène, puis que le traitement de la seconde scène en HF's de la séquence *coarse-to-fine* s'accompagnait d'une augmentation de l'activité du cortex occipital (en comparaison au traitement de la scène en BF's de la séquence *fine-to-coarse*) entre 170 et 200 ms après l'apparition de cette seconde scène. Plus important, l'augmentation de l'activité du cortex occipital lors du traitement de la scène en HF's ne survenait que lorsque cette scène était précédée d'une scène en

BFs, c'est-à-dire dans la séquence *coarse-to-fine*. Autrement dit, la comparaison de la source induite par la première scène en HFs de la séquence *fine-to-coarse* à celle induite par la première scène en BFs de la séquence *coarse-to-fine* ne montrait pas d'activation du cortex occipital. Les auteurs proposent qu'au cours de l'analyse *coarse-to-fine* d'une scène, les BFs seraient d'abord traitées par le cortex préfrontal, pour ensuite exercer des modulations descendantes sur le cortex occipital afin de guider le traitement ultérieur des HFs.

À noter que la région qui reçoit les prédictions dans cette étude serait donc le cortex occipital, et non pas le gyrus fusiforme comme dans l'étude de Bar et al. (2006). Cette interprétation s'inscrit alors plutôt dans le modèle intégratif de Bullier (2001) dans lequel le cortex visuel primaire est vu comme une région intégrative, codant à la fois les afférences rétino-thalamiques et les influences descendantes des aires corticales supérieures. Une étude en IRMf subséquente de cette même équipe (dont plusieurs résultats ont été publiés dans Musel et al. 2014 et Kauffmann et al. 2015b ; voir Kauffmann et al. 2015a; Musel et al. 2012 pour les études comportementales qui ont précédé ces travaux de neuroimagerie) a permis de conforter ces résultats. Pour rappel, dans cette étude, les participants devaient catégoriser des séquences composées d'images filtrées d'une même scène et assemblées en allant des BFs aux HFs (afin d'imposer une analyse *coarse-to-fine*) ou des HFs aux BFs (afin d'imposer une analyse *fine-to-coarse*) comme étant des scènes d'intérieur ou d'extérieur. Les analyses des données IRMf ont d'abord montré que les séquences *coarse-to-fine*, par rapport aux séquences *fine-to-coarse*, activaient le cortex frontal inférieur (au niveau du cortex orbitofrontal), le cortex occipito-temporal (au niveau du gyrus fusiforme et du cortex parahippocampique), et le cortex occipital (au niveau du cuneus, i.e. des aires visuelles de bas niveau). Sur la base de ces résultats, Musel et al. (2014) se sont particulièrement intéressés à la réponse de la PPA (*Parahippocampal Place Area*), une région du cortex parahippocampique fonctionnellement définie comme répondant plus fortement à des stimuli visuels de type scène qu'à des stimuli de type visage ou objet (Epstein 2005; Epstein & Kanwisher 1998). La PPA répondait plus fortement aux séquences *coarse-to-fine* qu'aux séquence *fine-to-coarse*, suggérant que cette stratégie d'analyse serait privilégiée pour le traitement visuel de scènes. À noter que Goffaux et al. (2010) ont également mis en évidence une prédominance de la stratégie d'analyse *coarse-to-fine* pour le traitement des visages au sein de la FFA (*Fusiform Face Area*), une région du cortex occipito-temporal cette fois-ci fonctionnellement définie comme répondant plus fortement à des stimuli visuels de type visage qu'à des stimuli de type scène ou objet (Kanwisher et al. 1997). Afin d'étudier spécifiquement les interactions entre les régions impliquées dans l'analyse *coarse-to-fine*, Kauffmann et al. (2015b) ont utilisé une méthode d'analyse de connectivité effective (*dynamic causal modeling* ; Friston, Harrison, & Penny 2003). Cette analyse a montré que le cortex frontal inférieur exerçait une influence descendante sur le cortex occipito-temporal (gyri fusiforme et parahippocampique) en accord avec les résultats de Bar et al. (2006) et Kveraga, Boshyan, et Bar (2007) mais aussi sur le cortex occipital (cuneus) en accord avec Peyrin et al. (2010).

Les résultats de cette équipe de recherche, et en particulier ceux de [Kauffmann et al. \(2015b\)](#), permettent d'appuyer le modèle proactif de Bar et de l'étendre à la reconnaissance de scènes en suggérant de nouveau que le cortex frontal inférieur, et plus particulièrement le cortex orbitofrontal, jouerait un rôle particulièrement important dans le codage prédictif de l'information visuelle. Cette structure corticale pourrait permettre d'emmagasiner rapidement l'information issue du *gist* de la scène et de la transformer en prédictions qui seraient ensuite renvoyées vers le cortex visuel primaire et les aires visuelles de la voie ventrale (comme le cortex inférotemporal) pour faciliter une analyse plus fine et minutieuse de la scène visuelle.

2.4. Modèle fréquentiel prédictif

Nous avons ici présenté deux modèles de reconnaissance visuelle, le modèle fréquentiel et le modèle proactif, et fait l'inventaire d'un certain nombre de résultats issus de la psychologie expérimentale, de la psychophysique, de la neurophysiologie et de la neuroimagerie qui corroborent ces modèles. Ces modèles sont interdépendants et sont basés sur le même postulat de base qui est celui d'une analyse *coarse-to-fine* de l'information visuelle, le modèle proactif spécifiant en plus une influence des BFs sur le traitement ultérieur des HFs, et une base neurale au modèle. Afin d'avoir un référentiel théorique clair sur lequel s'appuyer tout au long du manuscrit, nous proposons de formaliser un modèle qui synthétise les précédents en mettant l'accent sur leurs points communs : le modèle Fréquentiel Prédictif (Modèle FP ; Figure 1.13). Le raisonnement qui mène à ce modèle peut être résumé comme ceci :

- le système visuel est une sorte d'analyseur de Fourier qui décompose le signal en différentes fréquences spatiales ;
- l'analyse visuelle suivrait une analyse *coarse-to-fine* par défaut : les BFs sont traitées avec une précedence temporelle relativement aux HFs ;
- le cerveau comprend des aires corticales de haut niveau dans la hiérarchie qui répondent très rapidement à l'information visuelle, notamment grâce à des fibres magnocellulaires véhiculant une information en BFs ;
- par ailleurs, le cortex comprend de nombreuses connexions descendantes, permettant d'envisager l'existence d'influences descendantes des aires de haut niveau sur les aires de plus bas niveau ;
- les BFs pourraient ainsi permettre de faire rapidement des prédictions qui seraient ensuite utilisées pour guider le traitement plus lent des HFs.

Le modèle FP, que nous proposons de résumer en Figure 1.13, suppose donc que le système visuel décompose le signal comme un analyseur de Fourier. Une première partie de ce modèle correspond au modèle fréquentiel : Les BFs sont d'abord acheminées à un système de reconnaissance au temps 1 ($t1$) permettant l'émergence du *gist* suivies des HFs au temps 2 ($t2$). Une seconde partie du modèle FP correspond plutôt au modèle proactif, et plus généralement aux aspects de codage prédictif. Les BFs sont également envoyées, au $t1$, vers un système de gestion des prédictions et de l'erreur de prédiction.

Celui-ci envoie les prédictions vers le système de reconnaissance, qui accumule et intègre (i) les informations ascendantes provenant de la rétine, de manière *coarse-to-fine* et (ii) les informations descendantes (les prédictions) provenant du système de prédiction. Au niveau neurofonctionnel, la transmission des informations au $t1$ et au $t2$ serait gérée par les voies magnocellulaire et parvocellulaire, respectivement. Le système de reconnaissance correspondrait à la voie visuelle ventrale, depuis V1 jusqu'au cortex inféro-temporal, en passant par des aires visuelles de plus haut niveau comme la PPA ou le LOC. Le système de prédiction peut être implémenté dans le cortex orbitofrontal comme supposé par Bar (2003, Bar et al. 2006), ou de manière plus distribuée, à chaque étape du traitement visuel (ce qui correspondrait davantage à l'implémentation du modèle de Bullier 2001 ou encore à la conception de Rao & Ballard 1999).

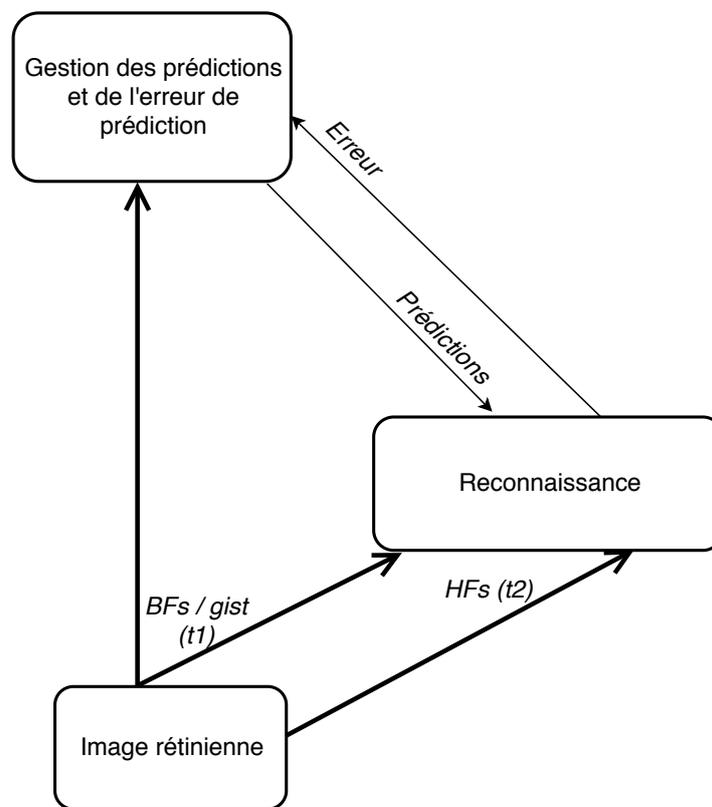


Figure 1.13: Modèle Fréquentiel Prédictif (modèle FP).

2.5. Importance d'un test direct de l'influence des BFs sur le traitement des HF

Le modèle FP fait donc clairement l'hypothèse d'une influence du contenu en BFs sur le traitement des HF. Pourtant, assez peu d'études ont, à notre connaissance, étudié directement l'influence du traitement des BFs sur le traitement ultérieur des HF, ou tout simplement l'intégration des différentes bandes de fréquences spatiales. Ces travaux seront présentés dans les paragraphes suivants. En effet, nous avons présenté jusqu'à présent deux types de résultats empiriques importants qui permettent d'appuyer le modèle

FP. Les premiers sont issus de travaux ayant étudié le rôle de l'ordre d'apparition des fréquences spatiales en comparant des séquences *coarse-to-fine* et *fine-to-coarse* (Kauffmann et al. 2015a, 2015b; Musel et al. 2012, 2014; Peyrin et al. 2010). Ces résultats apportent un argument important en faveur d'un modèle FP car ils montrent que le système visuel tire bénéfice d'avoir accès d'abord aux BFs puis aux HF. La seconde famille de résultats est issue de la neuroimagerie (e.g., Bar et al. 2006; Kauffmann et al. 2015c; Peyrin et al. 2010) et souligne le rôle de régions préfrontales, notamment du cortex orbitofrontal, et de régions occipito-temporales (e.g., cortex visuel primaire, cortex inférotemporal) dans le traitement des fréquences spatiales lors des processus de reconnaissance visuelle. Cependant, ces deux approches expérimentales ne permettent pas de tester directement l'influence du contenu en BFs sur le traitement des HF.

Cette problématique a été adressée par Mu et Li (2013) dans une étude en EEG, puis par Kauffmann et al. (2015c) dans une étude en IRMf, grâce à un paradigme expérimental d'interférence. Dans ces études, les stimuli étaient des scènes hybrides composées d'une scène en BFs superposée à une autre scène en HF (ces stimuli étant inspirés de l'expérience de Schyns & Oliva 1994). Les scènes appartenaient à deux catégories, "artificielle" (paysages urbains ou semi-urbains comme des rues de centre-ville, des autoroutes, etc.) et "naturelle" (paysages naturels de plage, de montagne, de forêt, etc.). Les deux scènes qui composaient une image hybride pouvaient être soit sémantiquement congruentes (e.g., les BFs d'une scène artificielle superposées aux HF d'une autre scène artificielle), soit sémantiquement incongruentes (e.g., les BFs d'une scène artificielle superposées aux HF d'une scène naturelle; Figure 1.14-A). Les auteurs attiraient l'attention des participants sur le fait que les stimuli contenaient deux scènes superposées, et leur demandaient de catégoriser (artificiel *vs* naturel) la scène en HF dans l'hybride, en ignorant le contenu en BFs. Dans le cadre d'un modèle FP, l'information en BFs serait automatiquement et rapidement traitée par les participants. Même si cette information n'est pas utile pour réaliser la tâche visuelle, elle influencerait la catégorisation de la scène en HF. Plus précisément, la prédiction opérationnelle que l'on peut faire est que les performances de catégorisation de la scène en HF devraient être moins bonnes si la scène en BFs est sémantiquement incongruente que si elle est sémantiquement congruente. On parle alors d'un effet d'interférence. Dans ces deux études, les auteurs ont également testé si l'effet d'interférence pouvait être modulé par la similarité physique entre les deux scènes de l'image hybride. Ainsi, dans chaque condition de congruence sémantique, les deux scènes des images hybrides étaient soit physiquement similaires (superposition spatiale - pixel par pixel - de l'information visuelle des deux scènes et similarité de leurs spectres d'amplitude), soit physiquement dissimilaires.

Les résultats comportementaux de ces études montrent tout d'abord un effet d'interférence sémantique. Les participants sont plus lents et font plus d'erreurs pour catégoriser la scène en HF de l'image hybride lorsqu'elle est sémantiquement incongruente à la scène en BFs que lorsqu'elle était sémantiquement congruente. Cet effet n'est pas sans rappeler l'effet d'interférence globale mis en évidence avec les stimuli hiérarchiques de

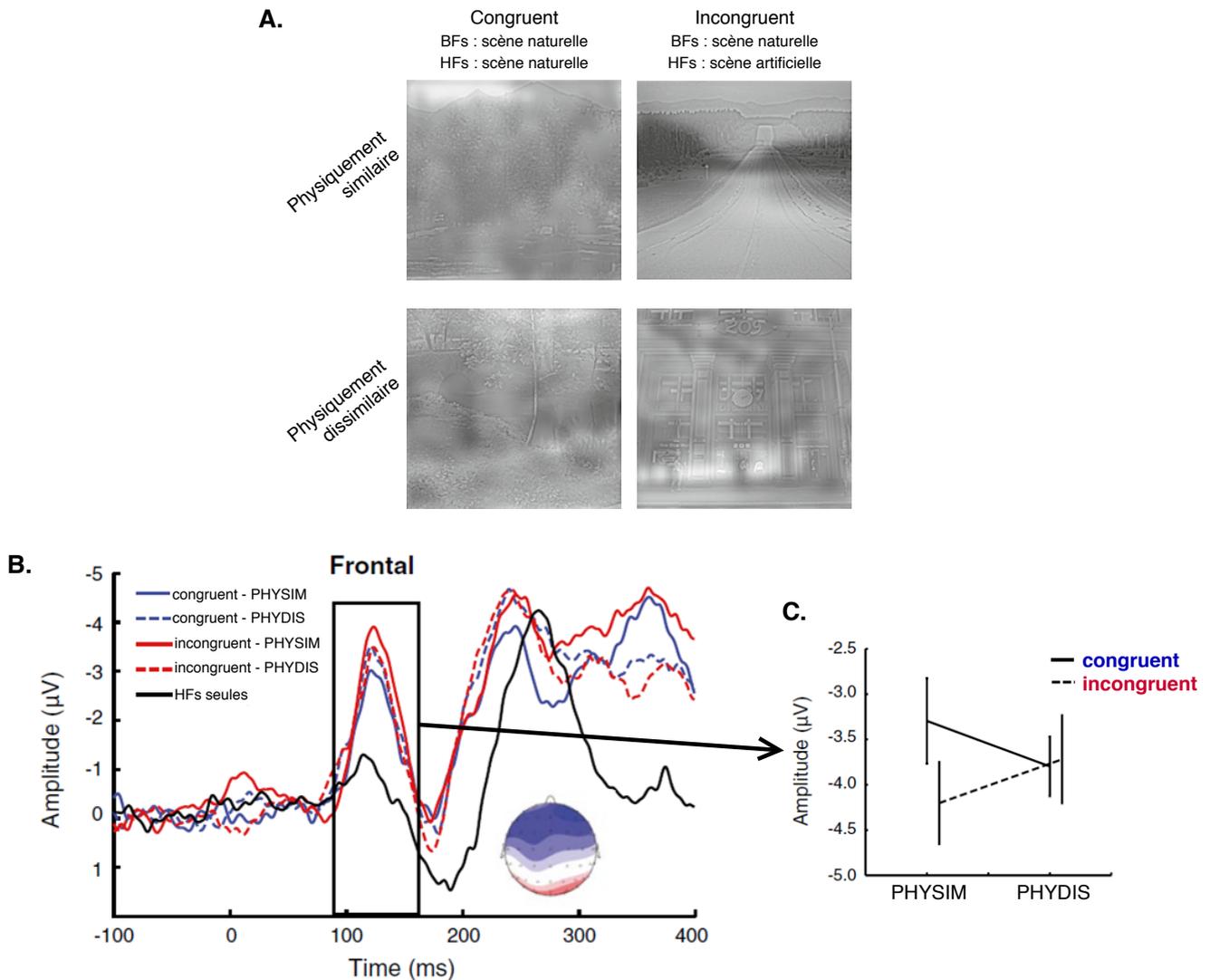


Figure 1.14: **A.** Exemples de stimuli utilisés dans l'étude de [Mu et Li \(2013\)](#). Les stimuli étaient des scènes hybrides constituées des BFs d'une scène superposées aux HF d'une autre scène. Les deux scènes composant un hybride pouvaient être sémantiquement congruentes ou incongruentes, et physiquement similaires (PHYSIM) ou physiquement dissimilaires (PHYDIS). **B.** Résultats de l'analyse des potentiels évoqués. La figure montre les courbes des potentiels évoqués moyens en fonction de la condition expérimentale, moyennés sur les électrodes frontales (Fz, F1, F2, FCz, FC1, FC2). La composante frontale précoce (N1) est encadrée en noir. **C.** Amplitude de la composante frontale N1 pour les différentes conditions expérimentales. Adapté de [Mu et Li \(2013\)](#).

[Navon \(1977\)](#). Concernant le rôle que joue la similarité physique entre les deux scènes, les résultats sont moins clairs. [Mu et Li \(2013\)](#) ont observé que l'effet d'interférence était encore plus important lorsque les deux scènes partageaient les mêmes propriétés physiques. Selon les auteurs, ce résultat suggère que l'intégration des BFs et des HF nécessite à la fois le traitement des propriétés physiques des scènes à un niveau perceptuel et un traitement sémantique à un niveau décisionnel. Cependant, ce résultat n'a pas été

répliqué par [Kauffmann et al. \(2015c\)](#).

Au niveau cérébral, les résultats en potentiels évoqués de [Mu et Li \(2013\)](#) ont montré que l'effet d'interférence sémantique était associé à une composante frontale précoce (N1) 122 ms seulement après l'apparition de l'image hybride (Figure 1.14-B et -C), suivie de composantes occipito-pariétale (P2) et occipitale (P3), respectivement 247 et 344 ms après l'apparition de l'image hybride. La différence d'amplitude entre les conditions congruente et incongruente sur la composante frontale précoce n'apparaissait que lorsque les deux scènes étaient physiquement similaires. Il est intéressant de noter que cette composante N1 frontale était en revanche d'amplitude bien plus faible lorsque les participants catégorisaient des scènes non hybrides en HFs dans une condition contrôle, montrant que c'était bien la présence de BF's qui modifiait l'activité des régions frontales de manière précoce. De façon cohérente avec les résultats de [Mu et Li \(2013\)](#), l'étude IRMf de [Kauffmann et al. \(2015c\)](#) a montré une activation plus importante du cortex frontal inférieur (au niveau du cortex orbitofrontal) et du cortex inférotemporal (au niveau du gyrus fusiforme et du gyrus parahippocampique) lorsque la scène en BF's était sémantiquement incongruente que sémantiquement congruente à la scène en HFs à catégoriser. L'activation du cortex orbitofrontal observée dans cette étude, cohérente avec les potentiels évoqués précoces sur les sites frontaux observés par [Mu et Li \(2013\)](#), suggère que des mécanismes d'intégration des BF's et des HFs pourraient être initiés par des régions préfrontales, et suggère également de nouveau le possible rôle de ces régions dans la gestion des prédictions issues du traitement des BF's. En outre, une analyse complémentaire de connectivité fonctionnelle a permis de montrer que l'effet d'interférence sémantique augmentait la connectivité fonctionnelle entre le gyrus fusiforme et le gyrus frontal inférieur.

Ainsi, les résultats de ces deux études montrent que l'information en BF's non pertinente pour la tâche serait traitée automatiquement et générerait la catégorisation de la scène en HFs. [Mu et Li \(2013\)](#) suggèrent que cette gêne serait d'autant plus forte lorsque l'agencement spatial des propriétés physiques des deux scènes, ainsi que leurs propriétés spectrales, sont similaires. Ces résultats fournissent des arguments supplémentaires en faveur d'un modèle FP en indiquant que l'information grossière contenue dans les BF's permet de former des prédictions sur le contenu possible de la scène. En cas d'incongruence sémantique, cette prédiction viendrait à l'encontre de l'information détaillée apportée par les HFs, ce qui donnerait lieu à un conflit sémantique très précoce (122 ms) impliquant les régions frontales et en une augmentation de la connectivité entre les régions frontales et occipito-temporales pour résoudre le conflit.

Plus récemment, [Petras, ten Oever, Jacobs, et Goffaux \(2019\)](#) ont étudié l'intégration et la contribution relative des BF's et des HFs au cours de la reconnaissance de visages en utilisant un tout autre paradigme expérimental. Les auteures ont entraîné un classifieur à discriminer des topographies de potentiels évoqués (enregistrés avec un EEG) par des stimuli dépourvus de sens en BF's (des visages *scrambled* dont le spectre de phase a été modifié par permutation aléatoire) de topographies évoquées par ce même type de

stimuli en HFs. Une fois que le classifieur "savait reconnaître" un pattern EEG lié soit au traitement de BFs, soit au traitement de HFs, il était alors utilisé pour labelliser le signal EEG (label BFs ou HFs) évoqué par des stimuli qui contenaient toutes les fréquences spatiales (i.e. stimuli tests). Ces stimuli tests pouvaient être (i) des visages intacts ou (ii) des visages *scrambled*, équivalents aux visages intacts en terme de spectre d'amplitude, mais non identifiables. Plus précisément, le classifieur était testé toutes les 10 ms sur le signal EEG enregistré à partir de la présentation du stimulus test. Ainsi, il était possible d'évaluer au cours du temps si le signal ressemblait plus à un traitement des BFs ou à celui des HFs. Les résultats ont montré qu'au début du traitement (notamment entre $\simeq 130-200$ ms après l'onset du stimulus), le classifieur attribuait significativement plus souvent le label BFs que HFs pour les topographies EEG évoquées par les visages intacts, ce qui n'était pas observé pour celles évoquées par les visages *scrambled*. Par contre, à partir de $\simeq 200$ ms, aucun label ne prédominait pour les visages intacts, alors que le classifieur attribuait cette fois significativement plus souvent le label HFs que BFs pour les visages *scrambled*. Autrement dit, les résultats montrent une baisse de la contribution des HFs dans le signal lorsqu'il y a une contribution forte des BFs au début, c'est-à-dire pour les visages intacts, mais pas pour les visages *scrambled*. Ce résultat suggère que lorsque le stimulus est porteur de sens (visages intacts), un traitement précoce des BFs permet de saisir une représentation robuste du visage qui permet ensuite de réduire le traitement (non nécessaire) des HFs. En ce sens, cette étude suggère une influence du traitement précoce des BFs sur le traitement des HFs.

Ainsi, les études de [Mu et Li \(2013\)](#), [Kauffmann et al. \(2015c\)](#) et [Petras et al. \(2019\)](#) ont testé directement l'influence du traitement des BFs sur les HFs et apportent ainsi des résultats qui nous semblent cruciaux pour soutenir un modèle FP. Dans le Chapitre 2, nous présenterons l'Expérience 1 de cette thèse, qui avait pour objectif de répliquer au niveau comportemental les résultats de [Mu et Li \(2013\)](#) et [Kauffmann et al. \(2015c\)](#), mais aussi de tester comment le traitement des HFs pouvait en retour influencer le traitement des BFs. En effet, il existe des preuves comportementales d'un traitement préférentiel des informations en HFs même pour des temps de présentation d'un stimulus très courts (e.g., [Özgen et al. 2005](#); [Rotshtein et al. 2010](#); [Schyns & Oliva 1999](#)). Cependant, le rôle de l'information en HFs et son impact sur le traitement des BFs est peu connu.

Pour conclure sur cette seconde partie introductive, nous avons vu que le modèle FP repose sur l'idée d'une dynamique de traitement *coarse-to-fine* de l'information visuelle. Ce modèle permet ainsi de rendre compte d'un grand nombre de résultats empiriques et plus largement de la rapidité et de l'efficacité de la reconnaissance visuelle. Toutefois, ce modèle en l'état actuel ne considère pas une particularité essentielle du système visuel humain, qui est qu'en pratique la résolution spatiale, et plus précisément la disponibilité de l'information en fréquences spatiales, n'est pas uniforme à travers le champ visuel. On entend par champ visuel la portion visible de l'environnement pour une image rétinienne donnée lors d'une fixation oculaire. Le champ visuel humain couvre environ $\simeq 200^\circ$ au total ($\simeq 100^\circ$ de chaque côté de la fixation; [Loschky, Szaffarczyk, Beugnet, Young, &](#)

Boucart 2019) dont environ $\simeq 120^\circ$ en vision binoculaire (le reste, c'est-à-dire les parties les plus excentrées du champ visuel, étant couvert seulement par l'extrémité de la rétine nasale d'un seul oeil). La résolution spatiale, ou acuité, est excellente dans une petite zone au niveau de la fixation. Cette zone mesure environ 1° d'angle visuel, ce qui correspond approximativement à la largeur du pouce lorsque nous tendons le bras. La région de la rétine sur laquelle se projette cette information est appelée la fovéa (Polyak 1941). La partie restante du champ visuel se projette (i) sur la parafovéa, qui couvre un diamètre d'environ 6° d'angle visuel centré sur la fovéa, et (ii) sur la rétine périphérique, qui couvre l'ensemble du champ visuel restant. Ainsi, alors que notre expérience visuelle subjective nous paraît riche et détaillée (en partie grâce à la constance des mouvements oculaires), la résolution spatiale diminue drastiquement avec l'excentricité et seule une information de basse résolution spatiale est disponible dans la majeure partie du champ visuel. Dans les faits (autrement dit, hors laboratoire), la majorité du signal en BFs provient donc de la vision périphérique. Dans la section qui suit, nous argumenterons pourquoi ce point devrait être considéré dans les modèles de reconnaissance visuelle. Notamment, nous ferons l'état d'arguments qui permettent d'établir une correspondance directe entre le concept de *gist* et le contenu visuel périphérique. À noter que lorsque nous évoquerons les régions du champ visuel dans la suite du manuscrit, nous distinguerons champ visuel central, qui comprend l'information lumineuse se projetant sur la fovéa et sur la parafovéa (approximativement un rayon de 3° autour du point de fixation), et champ visuel périphérique, qui couvre le reste du champ visuel. Cette distinction est plus ou moins arbitraire, puisqu'il n'existe pas de consensus scientifique sur la définition du champ visuel périphérique. D'ailleurs, la répartition non homogène des cellules de la rétine (Curcio, Sloan, Kalina, & Hendrickson 1990; Osterberg 1935) est telle que les propriétés d'extraction du signal au niveau de la rétine sont par exemple très différentes à 7° d'excentricité (où la majorité des photorécepteurs sont des cônes) et à 25° d'excentricité (où la majorité des photorécepteurs sont des bâtonnets). Cependant, ce qui est important pour notre problématique est de distinguer, même grossièrement, la portion du champ visuel autour de la fixation où l'extraction des HFs est possible versus le reste du champ visuel où seules les BFs sont disponibles.

3. Limite du modèle Fréquentiel Prédicatif : Qu'en est-il de la vision périphérique ?

3.1. Une correspondance directe entre le concept de *gist* et le contenu visuel périphérique

3.1.1. Résolution spatiale non homogène du champ visuel

Nous avons vu que les BFs sont transmises majoritairement par le biais de cellules M qui ont une grande vitesse de conduction, alors que les HFs sont transmises par le biais de cellules P qui répondent de manière plus lente. Au niveau rétinien, ces deux types de cellules ne sont pas répartis de manière équivalente selon si on se situe au niveau de la fovéa ou en périphérie de la rétine (Curcio & Allen 1990; Wässle et al. 1990). Au niveau de la fovéa, la densité des cellules ganglionnaires est grande et celles-ci reçoivent l'information provenant d'un nombre réduit de photorécepteurs, atteignant même, comme évoqué plus haut, un ratio d'une cellule ganglionnaire pour un photorécepteur au point central de la rétine. Autrement dit, le champ récepteur de ces cellules au centre de la rétine est très petit, et celles-ci sont en majorité des cellules P, à l'origine de la voie parvocellulaire. La densité de cellules ganglionnaires diminue ensuite avec l'excentricité rétinienne. Dans la périphérie de la rétine, les cellules ganglionnaires sont donc moins nombreuses et reçoivent l'information provenant d'un plus grand nombre de photorécepteurs (i.e. elles ont un grand champ récepteur). Ainsi, même si la densité des cellules P de la rétine ($\simeq 80\%$) est bien plus importante que celle des cellules M, on observe une augmentation de la densité des cellules M, à l'origine de la voie magnocellulaire, en périphérie de la rétine. Ces propriétés anatomiques de la rétine prédisposent à un traitement différencié de l'information lumineuse qui se projette sur la rétine fovéale par rapport à celle qui se projette sur la rétine périphérique. La rétine fovéale, grâce aux petits champs récepteurs des cellules ganglionnaires présentes dans cette région, peut traiter une information de bonne résolution spatiale alors que cette capacité diminue avec l'excentricité. Cela se traduit par le fait que théoriquement, l'extraction de HFs est seulement possible dans la partie centrale de la rétine, alors que l'information visuelle en BFs provient en grande majorité de la partie périphérique de la rétine (comme schématisé en Figure 1.15).

Une autre différence fondamentale entre la vision centrale et la vision périphérique concerne leur surface de représentation au niveau cortical. En effet, dès les corps genouillés latéraux, l'information visuelle n'est pas traitée de manière équivalente selon qu'elle provient du champ visuel central ou du champ visuel périphérique. Notons d'abord que chaque corps genouillé latéral traite l'information visuelle provenant de l'hémichamp visuel controlatéral. Ensuite, les cellules au sein de cette structure sont



Figure 1.15: Représentation schématique de la résolution spatiale non uniforme du champ visuel.

organisées de telle manière que des points proches du champ visuel sont codés par des cellules voisines dans les corps genouillés latéraux. Cette organisation "rétinotopique" permet de préserver l'organisation spatiale du champ visuel. Par ailleurs, comme dans la rétine, les cellules P de cette structure sont plus nombreuses que les cellules M. Cependant, ce ratio n'est pas le même qu'au niveau rétinien, et dépend de l'excentricité : le ratio P/M au niveau des corps genouillés latéraux est 3.9 fois supérieur au ratio P/M rétinien pour la région fovéale, alors qu'il est 2.7 fois inférieur au ratio P/M rétinien pour la périphérie (Figure 1.16 ; Azzopardi, Jones, & Cowey 1999), indiquant que la région centrale du champ visuel est sur-représentée par des cellules P au niveau des corps genouillés latéraux par rapport au champ visuel périphérique. Au niveau du cortex visuel primaire, l'organisation observée dans les corps genouillés latéraux est préservée, à la fois concernant la projection rétinotopique et la sur-représentation du champ visuel central. En effet, la surface corticale dédiée au traitement d'une région donnée du champ visuel diminue drastiquement avec l'excentricité, de telle sorte qu'une portion considérable de la surface du cortex visuel primaire est dédiée à la région fovéale (Qiu et al. 2006). Ce phénomène, que l'on appelle magnification corticale, est à nouveau à l'œuvre dans d'autres aires adjacentes au cortex visuel primaire, telles que V2 et V3.

Ainsi, dès la rétine et jusqu'au cortex visuel, l'information visuelle provenant du champ visuel périphérique est traitée différemment de l'information visuelle provenant du champ visuel central. Pour résumer, l'information en HF est extraite en vision centrale alors qu'une grande majorité de l'information en BF est extraite en vision périphérique. L'information périphérique se propage principalement par la voie magnocellulaire, tandis que l'information du champ visuel central se propage principalement par la voie

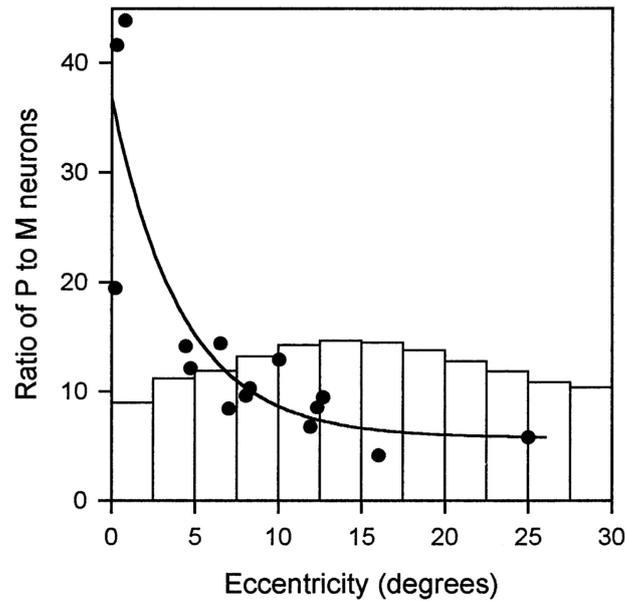


Figure 1.16: Ratio des cellules parvocellulaires (P) par rapport aux cellules magnocellulaires (M) au niveau des cellules ganglionnaires de la rétine (histogramme) et au niveau des corps genouillés latéraux (courbe), en fonction de l'excentricité (tiré de [Azzopardi et al. 1999](#)).

parvocellulaire. Les propriétés du système visuel semblent donc prédisposer à un traitement rapide et grossier (i.e. en BFs) de l'information périphérique et à un traitement plus lent de l'information détaillée (i.e. en HFs) extraite en vision centrale. A priori, et bien que les BFs peuvent aussi être extraites en vision centrale, il est donc clair que cette quantité considérable de signal en BFs issu de la périphérie doit avoir son importance dans le modèle FP.

3.1.2. Contribution de la vision périphérique dans la perception du *gist* d'une scène

En dehors du laboratoire, la perception globale de la scène engage la totalité du champ visuel et, par conséquent, fournit une grande quantité de signal en BFs rapidement accessible. On peut alors légitimement s'interroger sur le rôle que joue la vision périphérique dans la perception du *gist* d'une scène. À l'aide d'un écran panoramique couvrant 180° du champ visuel sur l'axe horizontal, [Boucart, Moroni, Thibaut, Szaffarczyk, et Greene \(2013\)](#) ont étudié la capacité à extraire les propriétés globales de scènes naturelles à des excentricités très larges. Dans une première expérience, deux scènes (20° × 20°) étaient présentées simultanément pendant 80 ms sur le méridien horizontal, l'une dans le champ visuel gauche et l'autre dans le champ visuel droit. À chaque essai, les participants devaient indiquer de quel côté se trouvait la scène qui présentait la propriété globale cible. Les propriétés globales étaient les mêmes que celles de l'étude de [Greene et Oliva \(2009b\)](#); voir sous-partie 1.3.2), c'est-à-dire l'ouverture, l'expansion, la naturalité (naturel *vs* urbain) et le fait qu'il s'agisse d'une scène d'intérieur (*vs* extérieur). Les

scènes étaient présentées à quatre excentricités (10° , 30° , 50° et 70°). Les résultats de cette étude sont assez remarquables, puisqu'ils indiquent que la reconnaissance de la plupart des propriétés globales des scènes est possible même à 70° d'excentricité, bien que cela dépende de la propriété globale à détecter. Par exemple, la précision pour déterminer le caractère naturel des scènes était supérieure à 90% jusqu'à 50° d'excentricité et supérieure à 70% à 70° d'excentricité. Par contre, la précision pour déterminer si une scène était un environnement d'extérieur ou d'intérieur était moins bonne : à peine à 80% de précision à 10° d'excentricité et au niveau de la chance à 70° . Une deuxième expérience de ces auteurs montre que la précision pour catégoriser les scènes au niveau de base (i.e. forêt, montagne, rivière, ville, autoroute, gratte-ciels) restait bonne ($>70\%$) jusqu'à 70° d'excentricité, et ce pour toutes les catégories testées.

Larson et Loschky (2009) ont directement testé la contribution relative de la vision centrale et de la vision périphérique dans la reconnaissance du *gist* des scènes. Ils présentaient de grandes scènes englobant le champ visuel central et le champ visuel périphérique proche ($27^\circ \times 27^\circ$) et utilisaient un paradigme *window-scotoma*. Les participants percevaient soit la scène complète (condition contrôle), soit seulement la portion centrale de la scène (condition *window*), soit seulement la portion périphérique (condition *scotoma*; Figure 1.17). La taille du rayon du scotome variait entre 1° et 13.6° . Les participants voyaient d'abord le stimulus (contrôle, *window* ou *scotoma*) présenté très brièvement ($\simeq 100$ ms), puis un label écrit désignant une catégorie niveau de base (e.g., "plage"). Ils devaient décider si le stimulus était apparié au label. Globalement, les résultats ont montré que les performances étaient meilleures lorsque l'information était disponible en vision périphérique que lorsqu'elle était disponible en vision centrale. Mais le résultat qui était peut être le plus marquant dans cette étude est que les performances dans la condition *scotoma* (donc en vision périphérique) étaient équivalentes à celles dans la condition contrôle jusqu'à une obstruction de la vision centrale d'un rayon d'environ 11° . Les résultats de Larson et Loschky (2009) ont récemment été répliqués par Geuzebroek et van den Berg (2018), mais aussi simulés à l'aide de réseaux de neurones convolutifs (Wang & Cottrell 2017). L'ensemble de ces résultats suggère que l'information provenant du champ visuel central et de la proche périphérie ne serait pas vraiment utile pour la reconnaissance rapide de scènes (alors même que, rappelons-le, le traitement de cette région du champ visuel occupe une vaste proportion de la surface corticale du cortex visuel primaire). Au contraire, l'information périphérique suffirait à saisir le *gist* de la scène.

Le contenu visuel périphérique serait donc d'une importance majeure dans l'extraction du *gist* des scènes et pourrait ainsi suffire à leur reconnaissance rapide. Néanmoins, les conclusions de l'étude de Larson et Loschky (2009) peuvent être limitées par le fait que la portion de la scène disponible dans la condition *window* ne contenait pas nécessairement d'information pertinente à la réalisation de la tâche (e.g., la portion visible en vision centrale pouvait comprendre seulement un fond uniforme correspondant à du sable dans une scène de plage). Dans un travail de collaboration avec l'équipe de recherche

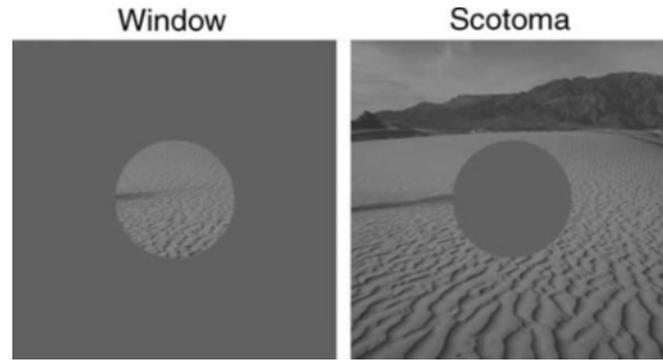


Figure 1.17: Paradigme expérimental *window-scotoma* (tiré de [Larson & Loschky 2009](#)).

du LPNC à laquelle j'appartiens, et dans le contexte de la thèse d'Audrey Trouilloud ([Trouilloud et al. soumis](#)), nous avons répliqué ces résultats en contrôlant l'information sémantique disponible en vision centrale. Les résultats de cette étude montrent que la catégorisation de scènes (intérieur *vs* extérieur) était meilleure en condition *scotoma* (vision périphérique) qu'en condition *window* (vision centrale) même lorsqu'un objet sémantiquement relié à la scène était contenu dans la portion centrale (condition *window*) de la scène (e.g., un bateau pour une scène de plage, un ustensile de cuisine pour une scène d'intérieur).

Pour résumer, ces études suggèrent qu'il est non seulement possible de reconnaître des scènes à de larges excentricités rétiniennes, notamment sur la base de leurs propriétés globales, mais aussi que la reconnaissance rapide de scènes est plutôt basée sur l'information périphérique que sur l'information centrale. Le contenu visuel périphérique serait donc d'une importance majeure dans l'extraction du *gist* des scènes et pourrait suffire à leur reconnaissance rapide.

3.1.3. *Crowding* et *summary statistics* en vision périphérique

La perception subjective d'un stimulus en vision périphérique n'est pas seulement contrainte par la baisse de résolution spatiale avec l'excentricité rétinienne. Un autre facteur qui détermine fortement la qualité perceptive (l'apparence) du stimulus en vision périphérique est le phénomène d'encombrement visuel, ou *crowding*, dans lequel la perception d'un objet donné en périphérie est détériorée en présence d'éléments adjacents. Un objet identifiable en isolation peut ainsi devenir totalement méconnaissable lorsqu'il est entouré d'autres objets (Figure 1.18). Le *crowding* suit la loi de Bouma, selon laquelle les éléments adjacents interfèrent avec la reconnaissance d'un objet en vision périphérique dès lors qu'ils sont situés à une distance égale ou inférieure à $e/2$ de l'objet, e étant la valeur d'excentricité de l'objet par rapport à la fixation centrale ([Bouma 1970](#)). La distance critique est donc proportionnelle à l'excentricité (i.e. plus l'objet cible est excentré, et plus il subira les influences d'éléments qui lui sont éloignés).

Le *crowding* se manifeste non seulement pour des stimuli simples comme des lettres

(Grainger, Tydgat, & Isselé 2010) ou des patchs de Gabor (Levi, Hariharan, & Klein 2002), mais aussi pour des stimuli plus complexes tels que des visages. Par exemple, il est plus difficile de reconnaître un visage dans la vision périphérique lorsque celui-ci est entouré d'autres visages que lorsqu'il est isolé (Farzin, Rivera, & Whitney 2009; Louie, Bressler, & Whitney 2007). Des résultats similaires ont aussi été observés pour la perception de mouvements biologiques (en utilisant des *point light displays*; Ikeda, Watanabe, & Cavanagh 2013) ou pour la reconnaissance d'objets (Kimchi & Pirkner 2015, pour une synthèse sur le phénomène de *crowding* en fonction des différents niveaux de complexité des stimuli, nous renvoyons à la récente revue de Manassi & Whitney 2018).



Figure 1.18: Deux exemples de *crowding*. Lorsqu'on fixe les croix rouges, il est facile de percevoir les formes situées dans le champ visuel gauche comme étant une barre orientée à 45° (A) ou comme étant la lettre G (B). Par contre, lorsque ces mêmes formes sont 'encombrées' comme c'est le cas dans le champ visuel droit, il devient difficile de juger de l'orientation de la barre en A ou d'identifier la lettre G en B.

Il est intéressant de noter que le *crowding* est en fait étroitement lié aux *summary statistics*. Par exemple, Parkes et al. (2001) ont montré que bien que le jugement de l'orientation d'un patch de Gabor est entravé lorsque celui-ci est entouré d'autres patchs de Gabor (i.e. *crowding*), il est possible de juger avec une bonne précision de l'orientation moyenne de l'ensemble de patchs, i.e. de former un percept d'ensemble. Un autre exemple concerne la perception d'ensembles de visages que nous avons déjà évoquée. Bien que la reconnaissance d'un visage donné soit rendue très difficile lorsqu'il est perçu au sein d'un ensemble de visages en périphérie (Farzin et al. 2009; Louie et al. 2007), il est possible d'extraire des informations moyennes d'un ensemble de visages, comme l'identité ou l'émotion (de Fockert & Wolfenstein 2009; Haberman & Whitney 2009). Une étude très récente (To, Carvey, Carvey, & Liu 2019) montre d'ailleurs que le phénomène de moyennage des expressions faciales est plus efficace lorsque les visages sont perçus en vision périphérique qu'en vision fovéale, soulignant les liens étroits qui existeraient entre

les *summary statistics* et le *crowding* en vision périphérique. Ces résultats suggèrent que les éléments encombrés ne sont pas réellement perdus en vision périphérique mais qu'ils sont plutôt combinés au signal avoisinant.

Des études ont d'ailleurs montré que des informations subissant les effets du *crowding* en périphérie peuvent malgré tout influencer des jugements subséquents de manière non consciente (Atas, Faivre, Timmermans, Cleeremans, & Kouider 2013; Kouider, Berthet, & Faivre 2011). Par exemple, Kouider et al. (2011) présentaient à des participants des visages exprimant soit la colère, soit la joie. Ces visages étaient entourés de patterns texturés dénués de sens, de façon à ce que l'expression faciale ne puisse pas être consciemment accessible (puisque encombrée). Immédiatement après, les participants devaient juger du caractère agréable ou désagréable d'un pictogramme chinois (ces participants ne connaissant pas le chinois, les pictogrammes étaient donc considérés comme des stimuli neutres). Les résultats ont montré que les pictogrammes étaient jugés plus agréables lorsqu'ils étaient amorcés par un visage exprimant la joie plutôt que par un visage exprimant la colère, indiquant que l'émotion exprimée par le visage, bien que non consciemment perçue à cause du *crowding*, influençait la valence perçue d'un stimulus neutre. Dans la même idée, une étude de Yeh, He, et Cavanagh (2012) a montré que l'effet d'amorçage d'un mot présenté en condition de *crowding* sur une tâche de décision lexicale est aussi fort que l'effet d'amorçage d'un mot présenté en isolation. Ainsi, le *crowding* pourrait permettre de condenser l'information dans des percepts d'ensemble cohérents et pertinents, au sein desquels l'information n'est pas totalement perdue mais peut être au contraire traitée inconsciemment sous forme de *summary statistics* des diverses propriétés visuelles de l'environnement.

Plus haut dans ce chapitre, nous avons proposé l'idée que les BFs et les *summary statistics* peuvent conjointement constituer une bonne description du contenu perceptuel disponible lors de la perception du *gist*. Dans cette sous-partie, nous avons vu (i) que la majorité du signal en BFs provient de la vision périphérique et (ii) que les *summary statistics* seraient un mécanisme plutôt lié à la vision périphérique qu'à la vision centrale, notamment parce que les *summary statistics* reposeraient sur le phénomène de *crowding*. Ces deux points nous permettent de défendre ici l'idée d'une correspondance étroite entre la vision périphérique et la perception du *gist*. Nous pensons ainsi que la perception du *gist* serait plus efficace si elle s'appuie sur la vision périphérique que sur la vision centrale (Larson & Loschky 2009) parce que la perception d'une scène en vision centrale limiterait la possibilité (i) d'extraire l'information en BFs et (ii) d'utiliser des *summary statistics*.

Si les BFs et plus généralement le *gist* sont extraits majoritairement en vision périphérique, alors un modèle FP prédit une influence des BFs périphériques sur le traitement des HFs en vision centrale. Cette hypothèse est la base du travail empirique qui a été réalisé au cours de cette thèse. Dans la partie suivante, nous présenterons des résultats de recherche qui, sans avoir été directement motivés par cette hypothèse, corroborent cette idée.

3.2. Arguments en faveur de l'hypothèse d'une influence de la vision périphérique sur la vision centrale

Une façon d'envisager les différences qualitatives qui existent entre la vision centrale et la vision périphérique est de considérer le type d'informations sémantiques auxquelles elles sont généralement associées. Ces différences peuvent être modélisées par une distinction objet/contexte. En effet, les objets d'une scène visuelle sont rarement perçus en isolation, et sont le plus souvent rencontrés au sein d'un contexte spatial typique (e.g., une casserole a plus de chances d'être rencontrée dans le contexte de la cuisine que dans un autre contexte). Le système pourrait donc extraire des régularités statistiques permettant d'estimer les probabilités d'observer des objets dans un contexte donné (Bar & Ullman 1996). Ainsi, la reconnaissance d'une partie de la scène pourrait permettre de faciliter la reconnaissance d'objets adjacents. Cette idée a en fait été l'objet de nombreuses études qui se sont intéressées aux interactions visuelles entre objets et contexte (voir, entre autres, Bar & Ullman 1996; Biederman 1972; Boucart et al. 2013; Brandman & Peelen 2017; Davenport 2007; Davenport & Potter 2004; Joubert, Fize, Rousselet, & Fabre-Thorpe 2008; Joubert et al. 2007; Katti, Peelen, & Arun 2017; Mudrik, Lamy, & Deouell 2010; Mudrik, Shalgi, Lamy, & Deouell 2014; Palmer 1975; Rémy et al. 2013; Sun et al. 2016; Truman & Mudrik 2018). Par exemple, Palmer (1975) a étudié la manière dont un contexte pertinent pouvait influencer l'identification d'un objet. Les participants de cette étude devaient identifier des dessins d'objets qui étaient précédés par un contexte (e.g., une cuisine) présentés de manière rapide. L'objet était soit congruent avec le contexte (e.g., une niche de pain), soit incongruent (e.g., une boîte aux lettres). La probabilité d'identifier correctement l'objet était plus forte lorsque le contexte était congruent. Des études plus récentes ont étudié les effets de congruence entre un objet et son contexte à l'aide de photographies de scènes (Davenport 2007; Davenport & Potter 2004; Joubert et al. 2008, voir Figure 1.19). Dans ces études, les participants devaient catégoriser une image d'un objet (e.g., un tracteur), alors que celui-ci apparaissait sur une large photographie d'une scène représentant un contexte congruent (e.g., une ferme) ou incongruent (e.g., une banquise). De nouveau, les résultats montrent un effet de facilitation de la catégorisation de l'objet lorsque le contexte est congruent. Cet effet facilitateur du contexte congruent a également été observée avec le même type de paradigme chez des patients atteints de dégénérescence maculaire liée à l'âge (DMLA), une maladie rétinienne qui touche particulièrement la macula et se traduit donc par une perte progressive de la vision centrale (Boucart et al. 2013). Dans cette étude, les patients atteints de DMLA, tout comme des sujets contrôles appariés en âge, avaient de meilleures performances pour catégoriser des objets (meubles *vs* animaux) lorsqu'ils étaient insérés sur une scène représentant un contexte congruent (e.g., un salon pour une table) qu'incongruent (e.g., un champ pour une table). Ce résultat suggère ainsi que la perturbation de la reconnaissance d'objet liée à un déficit du champ visuel central des patients DMLA pourrait être partiellement compensée par une information contextuelle pertinente. Les auteurs ont d'ailleurs proposé que le contexte de la scène avait été traité

3. Limite du modèle Fréquentiel Prédicatif : Qu'en est-il de la vision périphérique ?

de manière implicite par les patients et a été utilisé pour faire des prédictions sur la nature des objets pouvant être trouvé dans ce contexte.

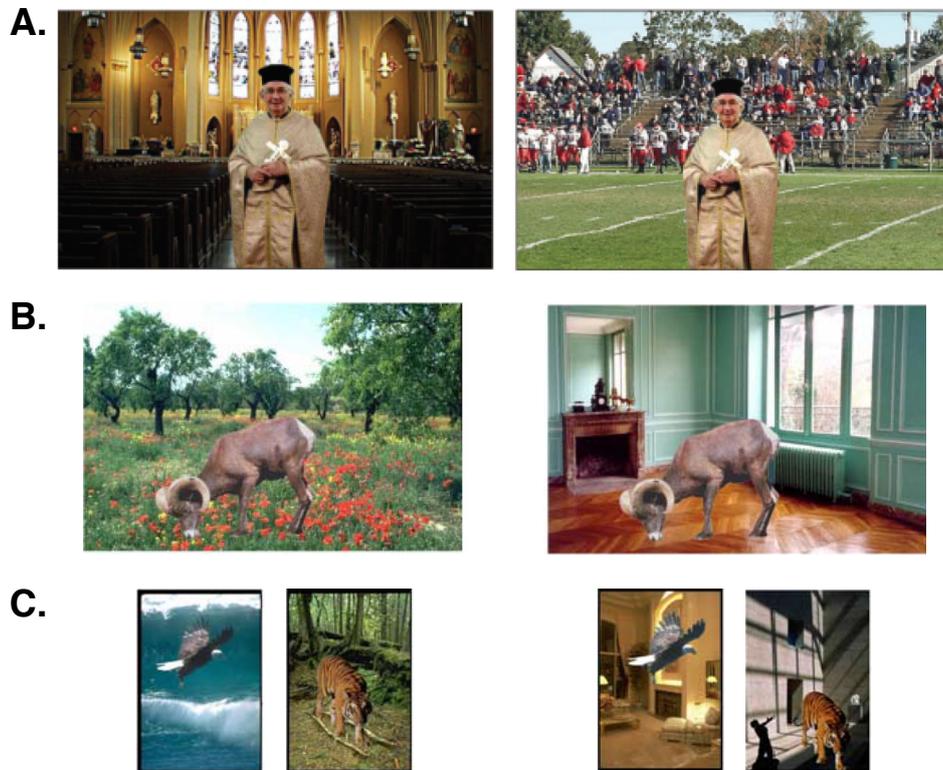


Figure 1.19: Exemples de stimuli utilisés dans les études de l'effet du contexte sur la reconnaissance d'objets. **A.** Davenport et Potter (2004). **B.** Boucart et al. (2013). **C.** Joubert et al. (2008). L'effet de contexte ou de congruence se manifeste par de meilleures performances pour catégoriser les objets lorsque le contexte visuel est congruent (stimuli de gauche) que lorsqu'il est incongruent (stimuli de droite).

Mais les mécanismes qui régissent ces effets de contexte sont assez peu connus. Selon Henderson et Hollingworth (1999), on peut différencier au moins deux hypothèses qui permettent d'expliquer les mécanismes d'identification des objets au sein d'une scène/d'un contexte selon le niveau de traitement auquel la scène exercerait une influence sur le traitement de l'objet. La première suppose que le contexte influence le traitement de l'objet à un niveau perceptif *via* des mécanismes de rétroaction (descendants). Cette vue est compatible avec le modèle FP. Dans ce cadre, l'information contextuelle serait captée rapidement grâce au traitement de l'information en BFs et pourrait avoir une influence prédictive, *via* des influences descendantes, sur la construction de la représentation perceptuelle de l'objet. La seconde hypothèse suppose au contraire que le contexte et l'objet interagissent à un niveau sémantique. Ils seraient alors traités de manière principalement indépendante, en parallèle, et leur intégration aurait lieu seulement après leur traitement perceptif, à un niveau sémantique. Les auteurs suggèrent même que selon cette hypothèse, le contexte de la scène pourrait influencer non pas l'identification de l'objet mais les processus de décision qui mènent à la réponse comportementale. Il existe

des arguments empiriques pouvant défendre cette perspective. Par exemple, nous avons déjà évoqué le fait qu'il existe des régions fonctionnelles dans le cortex occipito-temporal qui sont préférentiellement impliquées dans le traitement visuel des objets d'une part (*lateral occipital complex*, LOC) et des scènes d'autre part (*parahippocampal place area*, PPA). Les représentations visuelles des scènes et des objets pourraient ainsi exister de manière indépendante au niveau cortical.

L'approche adoptée par [Brandman et Peelen \(2017\)](#) dans une étude combinant mesures d'IRMf et de MEG pour dissocier les deux hypothèses (influence descendante du contexte sur l'objet *vs* traitement parallèle) était de pixeliser l'objet, rendant sa reconnaissance quasiment impossible hors contexte. Les résultats comportementaux ont tout d'abord montré que la reconnaissance de ces objets pixellisés était bien meilleure lorsqu'ils étaient intégrés dans des scènes congruentes que lorsqu'ils étaient isolés (i.e. présentés sur un fond gris). Des analyses multivariées du signal BOLD dans le LOC ont ensuite montré que les représentations de l'objet dans cette région sont plus facilement décodables par un classifieur lorsqu'un contexte congruent est présent. Les résultats MEG indiquent que l'effet facilitateur du contexte émerge à partir de 320 ms après l'onset du stimulus, soit environ 100 ms plus tard que le temps habituellement nécessaire pour reconnaître des objets intacts ([Carlson, Tovar, Alink, & Kriegeskorte 2013](#)). Selon les auteurs, il est peu probable qu'un tel délai reflète un traitement purement ascendant. Ce temps de traitement supplémentaire refléterait selon eux les processus nécessaires au traitement du contexte et à la rétroinjection de ces informations sur le cortex visuel extrastryé (comme le LOC) pour faciliter le traitement perceptuel de l'objet pixellisé.

Ces résultats suggèrent alors que les mécanismes de traitement du contexte et de l'objet ne sont pas indépendants, et que l'information contextuelle aide la catégorisation de l'objet en modulant sa représentation perceptive. Ainsi, la reconnaissance d'un objet dont la qualité perceptive est très mauvaise pourrait bénéficier d'influences issues du traitement du contexte. En revanche, de notre point de vue, il est possible que ces interactions entre objet et contexte puissent se produire de manière parallèle, sans nécessairement recourir à des mécanismes de rétroaction. Par exemple, le temps de traitement supplémentaire pour traiter un objet pixelisé dans un contexte congruent pourrait simplement traduire les mécanismes de reconstruction du percept (initialement dégradé) de l'objet grâce à des interactions horizontales avec l'information contextuelle. Dans le Chapitre 2, nous proposons de nouvelles approches pour tester l'hypothèse d'une influence prédictive d'une scène présentée en vision périphérique sur la reconnaissance d'objet en vision centrale (Expériences 4 et 5).

Enfin, un dernier argument en faveur de l'hypothèse d'une influence de la vision périphérique sur la vision centrale provient d'une expérience en IRMf de [Williams et al. \(2008\)](#), où les auteurs ont observé de façon totalement inattendue que des aires visuelles rétiniotopiques dédiées à la vision centrale pourraient coder l'information visuelle périphérique. Leur expérience était initialement programmée pour étudier si les représentations d'objets dans le cortex visuel sont invariantes à la position. Les participants fixaient le

centre de l'écran, et des objets apparaissaient en vision périphérique. À chaque essai, deux objets apparaissaient simultanément dans des quadrants périphériques opposés sur la diagonale (e.g., un objet en haut à droite et un objet en bas à gauche). Les auteurs ont réalisé des analyses de corrélation entre les patterns spatiaux d'activations (*multivariate pattern analysis*) en fonction de leurs conditions expérimentales d'intérêt. Ces conditions expérimentales ainsi que les hypothèses initiales des auteurs n'ont pas grand intérêt à être détaillées ici. Le résultat surprenant et qui se révèle particulièrement intéressant pour notre problématique de recherche, est que dans le cortex rétinotopique fovéal, les patterns d'activation pour des objets de même catégorie corrélaient. En d'autres termes, l'activité dans le cortex rétinotopique fovéal pouvait permettre de retrouver la catégorie des objets présentés en vision périphérique, alors même qu'aucun stimulus n'était présenté en vision centrale. De plus, cette information codée dans le cortex rétinotopique fovéal était indépendante de la position de l'objet dans le champ visuel périphérique, indiquant qu'elle contenait une information de relativement 'haut niveau' (non rétinotopique). Ce phénomène remet singulièrement en question ce qui est attendu et connu du cortex rétinotopique.

L'expérience n'ayant pas été initialement prévue pour mettre en évidence ce genre d'effet, les auteurs ont réalisé cinq répliques indépendantes. Par ailleurs, à l'aide de différents contrôles, ils ont montré que ce résultat n'est ni imputable aux mouvements oculaires, ni à des activations liées à la croix de fixation présentée fovéalement, ni non plus à des activations rétinotopiques périphériques qui déborderaient sur le cortex rétinotopique fovéal. Comme dit plus haut, ces résultats sont importants à considérer dans le cadre de nos recherches puisqu'ils suggèrent que des stimuli périphériques peuvent modifier l'activité du cortex rétinotopique fovéal. Par ailleurs, l'information d'origine périphérique codée dans le cortex rétinotopique fovéal pourrait être une information de relativement haut niveau (i.e. invariante à la position). Or, et c'est un point important, un modèle de reconnaissance visuelle qui suppose que l'information visuelle périphérique est prédictive de l'information visuelle centrale requiert l'existence de mécanismes qui permettent la rétro-injection de prédictions non rétinotopiques .

4. Objectif des travaux de thèse et méthodologie générale

4.1. Paramètres expérimentaux généralement adoptés

Nous avons répertorié dans le Tableau 1.1 les paramètres expérimentaux des expériences qui confortent le modèle FP et/ou les effets de contexte, lorsque ceux-ci sont disponibles. Il s'agit d'une revue non exhaustive mais qui permet néanmoins d'avoir une vue d'ensemble sur le type de stimuli et la manière dont ceux-ci sont présentés dans les expériences importantes pour le modèle. On peut constater que parmi elles, une large majorité a effectivement utilisé des petits stimuli présentés dans la vision centrale des participants. On peut attribuer cette manière de procéder à des raisons qui sont le plus souvent liées aux conditions d'expérimentation en laboratoire, aux traditions de la psychologie expérimentale et de la psychophysique et aussi au simple fait qu'il paraît naturel de présenter les stimuli "là où on voit le mieux". Il y a cependant deux exceptions notables quant à la taille des stimuli. La première est celle des expériences évaluant le rôle de l'ordre d'apparition des fréquences spatiales (séquences *coarse-to-fine* et *fine-to-coarse*; Kauffmann et al. 2015a, 2015b; Musel et al. 2012, 2014). Ces études utilisaient des larges scènes (généralement $24^\circ \times 18^\circ$) qui couvraient à la fois la vision centrale et une partie de la vision périphérique. Cependant, le filtrage en fréquences spatiales (en BFs ou HFs) était appliqué sur toute la surface du stimulus, ce qui ne permet donc pas de savoir quelle était la contribution de la vision centrale et de la vision périphérique dans ce paradigme. La deuxième exception est celle de trois expériences évaluant l'effet du contexte sur la reconnaissance d'objet (Boucart et al. 2013; Davenport 2007; Davenport & Potter 2004, voir Figure 1.19-A et -B). Toutefois dans ces études, la position de l'objet sur la rétine était incertaine. L'objet n'était pas nécessairement perçu en vision centrale et sa taille n'était pas contrôlée. La position et la taille de l'objet n'étaient d'ailleurs pas non plus contrôlées dans les autres études évaluant l'effet de contexte (e.g., Bar & Ullman 1996; Biederman 1972; Brandman & Peelen 2017). Autrement dit, jusqu'ici, ces études se sont intéressées à l'effet du contexte sur la reconnaissance d'objet sans considérer les relations objet-vision centrale et scène-vision périphérique.

4.2. Objectif : Test d'un modèle FP-peri

Le modèle FP prévoit que les BFs sont traitées très rapidement, que cette première analyse grossière permet d'activer des prédictions, et que ces prédictions sont ensuite utilisées afin de guider le traitement subséquent des HFs. Par extension, le modèle FP-peri (pour modèle Fréquentiel Prédicatif intégrant la vision périphérique; voir Figure 1.20) prévoit que les BFs extraites en vision périphérique devraient jouer le même rôle :

Tableau 1.1: Revue non exhaustive des paramètres expérimentaux adoptés dans des études qui permettent d'étayer le modèle FP (taille du stimulus et position du stimulus dans le champ visuel, en fonction du type d'effet mis en évidence). CV : champ visuel

Référence	Taille des stimuli	Position dans le CV	Effet mis en évidence
Badcock et al. (1990)	$1.96^\circ \times 3.31^\circ$	CV central	précédence globale et influence des fréquences spatiales
Bar et Ullman (1996)	$9.5^\circ \times 6.5^\circ$ (objets)	CV central + proche périphérie	effet de contexte (intra-objet)
Bar et al. (2006)	4.5° sur le côté le plus large	CV central	influence prédictive basée sur les BFs
Biederman (1972)	$5^\circ \times 3.5^\circ$ (scène)	CV central, position incertaine de l'objet dans le CV	effet de contexte
Boucart et al. (2013)	$20 \times 15^\circ$	position incertaine de l'objet dans le CV	effet de contexte
Bonnar et al. (2002)	$5.72^\circ \times 5.72^\circ$	CV central	global/local
Brandman et Peelen (2017)	$\sim 6^\circ \times 5^\circ$	position incertaine de l'objet dans le CV	effet de contexte
Davenport (2007)	$22 \times 13^\circ$	position incertaine de l'objet dans le CV	effet de contexte
Davenport et Potter (2004)	$22 \times 13^\circ$	position incertaine de l'objet dans le CV	effet de contexte
Goffaux et al. (2010)	$5.8^\circ \times 5.8^\circ$	CV central	<i>coarse-to-fine</i> (visages)
Kauffmann et al. (2015a)	$24^\circ \times 18^\circ$	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Kauffmann et al. (2015)	$24^\circ \times 18^\circ$	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Kauffmann et al. (2015b)	$24^\circ \times 18^\circ$	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Mu et Li (2013)	$7.2^\circ \times 7.2^\circ$	CV central	interférence des BFs sur les HF
Musel et al. (2012)	$24^\circ \times 18^\circ$	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Musel et al. (2014)	16.4×16.4 degrees	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Parker et Salzen (1977)	6° (diamètre)	CV central	<i>coarse-to-fine</i> (réseaux sinusoidaux)
Petras et al. (2019)	$5.5^\circ \times 4.5^\circ$	CV central	intégration des BFs et HF
Peyrin et al. (2010)	$4^\circ \times 4^\circ$	CV central	<i>coarse-to-fine</i> (scènes)
Poirel et al. (2010, 2008)	$8.5^\circ \times 11.4^\circ$	CV central + proche périphérie	précédence globale
Ramanoël et al. (2015)	$24^\circ \times 18^\circ$	CV central + périphérique	<i>coarse-to-fine</i> (scènes)
Rotshtein et al. (2010)	$3.5^\circ \times 3.5^\circ$	CV central	<i>coarse-to-fine</i> (objets)
Schyns et Oliva (1994)	$6.27^\circ \times 4.38^\circ$	CV central	<i>coarse-to-fine</i> (scènes)
Shulman et al. (1986)	$1^\circ \times 1^\circ$	CV foveal	précédence globale et influence des fréquences spatiales

permettre d'activer des prédictions qui sont ensuite utilisées pour guider le traitement des HFs (nécessairement) extraites en vision centrale. Par rapport au modèle FP initial, les différences fondamentales sont donc que **(i)** le système de reconnaissance reçoit des inputs en BFs provenant de la rétine centrale et de la rétine périphérique, et des inputs en HFs provenant uniquement de la rétine centrale, et **(ii)** le système de gestion des prédictions reçoit des inputs en BFs provenant de la rétine centrale mais surtout de la rétine périphérique (cf. Figure 1.20). L'objectif principal des travaux expérimentaux réalisés au cours de cette thèse a été de tester l'hypothèse du modèle FP-péri, et d'apporter des éléments de réponse à la question plus large qui serait comment les informations contextuelles perçues en vision périphérique peuvent influencer la reconnaissance visuelle en vision centrale? Afin de tester notre hypothèse, nous avons utilisé deux approches : **(i)** des tâches de catégorisation dans des paradigmes expérimentaux d'interférence, dans la lignée de ce qui a été fait dans les études de [Mu et Li \(2013\)](#) et [Kauffmann et al. \(2015c\)](#), afin de tester l'existence d'une interférence de l'information périphérique sur les traitements réalisés en vision centrale, et **(ii)** des tâches de catégorisation de stimuli présentés dans la vision centrale de patients atteints d'un glaucome, afin d'évaluer l'impact d'une vision périphérique déficitaire sur les processus résiduels de reconnaissance visuelle en vision centrale. Nous détaillons dans la partie suivante les raisons de ces choix méthodologiques.

4.3. Tâches de catégorisation

Au début de ce document, nous avons proposé une définition de la reconnaissance visuelle : la correspondance entre un contenu perceptif et une représentation interne porteuse de sens, qui donne lieu à la sensation de pouvoir attribuer un sens à ce qui nous entoure. Nous avons évoqué le fait que cette notion de reconnaissance visuelle était inextricablement liée au principe de catégorisation cognitive. La catégorisation peut être comprise comme la manière dont nous regroupons les choses qui se ressemblent. Vu la diversité des choses dans notre environnement, nous ne sommes pas capables d'encoder une représentation parfaite de chaque exemplaire de chaque catégorie d'objet. De tels mécanismes seraient considérablement coûteux computationnellement. Au contraire, le système de mémoire excelle dans le fait d'extraire les similarités et les différences existantes entre différents concepts. Il forme donc des catégories. Comme nous l'avons vu, des travaux en psychologie expérimentale suggèrent que le niveau de base serait le niveau de catégorisation favorisé, justement parce qu'il trouve le meilleur compromis entre l'exigence d'une grande similarité des exemplaires au sein de la catégorie et celle d'une suffisante distance d'avec d'autres catégories.

Le principe de catégorisation est certes intuitif, mais il a aussi été corroboré par des résultats de recherche. Le fameux principe des "neurones grand-mère" en est une bonne illustration. Ce principe suppose l'existence d'unités de reconnaissance qui sont invariantes à la position et aux changements structurels (e.g., un neurone qui répondrait spécifiquement à des stimulations impliquant ma grand-mère, quel que soit son angle de

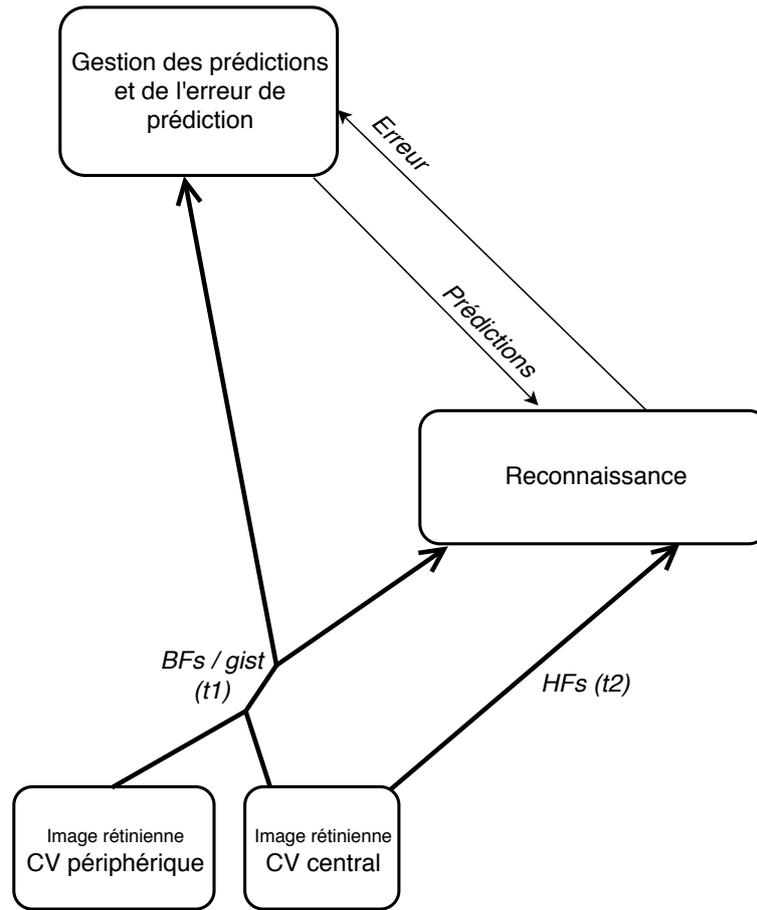


Figure 1.20: Modèle Fréquentiel Prédicatif intégrant la vision périphérique (modèle FP-peri). CV : champ visuel.

vue, son expression faciale ou encore sa coiffure). L'hypothèse des neurones grand-mère est caricaturale et représente davantage un cadre de pensée qu'une réalité cognitive et biologique (Gross 2002). Cependant, il existe bel et bien des neurones du cortex inféro-temporal qui ont de larges champs récepteurs et qui peuvent ainsi répondre à des stimuli complexes avec une résistance au point de vue (Booth & Rolls 1998; Tanaka 1993). Cette capacité du cerveau à généraliser des stimuli visuels semblables lui permet donc de catégoriser en évitant de devoir recourir à des représentations trop précises. Ainsi, la reconnaissance visuelle (rapide) est possible en faisant correspondre un percept unique (puisque l'on peut supposer qu'il est impossible d'avoir exactement la même stimulation visuelle deux fois) à une catégorie existante en mémoire.

Pour cette raison, la manipulation expérimentale des catégories dans des tâches de catégorisation nous paraît être une bonne méthode pour évaluer la reconnaissance visuelle. Le postulat simple des tâches de catégorisation est que si un stimulus est catégorisé correctement, alors il a été reconnu. Utiliser des tâches de catégorisation nécessite bien évidemment d'utiliser des stimuli visuels impliquant des représentations de 'haut niveau' (qui ont une signification et peuvent donc être reconnus : scènes, objets, visages). Nous avons donc utilisé des photographies de scènes naturelles, d'objets et de visages.

Nous avons pris le parti de privilégier la validité écologique de nos stimuli en utilisant des stimuli plutôt naturels, mais qui par conséquent limitent le contrôle des propriétés de bas niveau au niveau de l'image. Nous avons tout de même limité les paramètres non contrôlés en utilisant seulement des stimuli en niveau de gris et en contrôlant certains paramètres de nos images comme la luminance moyenne ou le contraste.

4.4. Approche méthodologique chez le sujet jeune sain :

Le paradigme d'interférence sémantique

Alors qu'un sujet réalise une tâche de catégorisation, il est possible de manipuler une infinité de variables en fonction de la problématique en jeu. Dans notre cas, nous avons utilisé le paradigme d'interférence sémantique² en faisant varier systématiquement l'information présentée en vision périphérique alors que le sujet catégorise des stimuli présentés en vision centrale. En manipulant l'information sémantique présente en vision périphérique (notamment en faisant varier sa congruence avec le stimulus en vision centrale), il est ainsi possible de calculer un effet d'interférence, qui peut être considéré comme étant la résultante d'une intégration des deux informations (périphérique et centrale) en un tout. L'effet d'interférence peut ensuite être modulé par d'autres variables, notamment des variables davantage liées aux caractéristiques de bas niveau des images.

Les travaux réalisés à l'aide de ce paradigme chez le sujet sain sont présentés dans le Chapitre 2 (Expériences 1 à 5) :

- L'**Expérience 1** est en fait un prémisses aux suivantes. Dans cette expérience, nous avons utilisé le paradigme d'interférence avec des scènes hybrides présentées uniquement en vision centrale. Nous avons pour objectif **(i)** de répliquer les résultats obtenus par [Mu et Li \(2013\)](#) et [Kauffmann et al. \(2015c\)](#), à savoir une interférence des BFs lors de la catégorisation de HFs, **(ii)** de clarifier le rôle de la similarité physique des scènes dans cet effet, puisque les résultats de [Mu et Li \(2013\)](#) et [Kauffmann et al. \(2015c\)](#) ne sont pas clairs à ce sujet, et, **(iii)** de tester comment les HFs peuvent en retour influencer la perception et la catégorisation de la scène en BFs.
- Dans les **Expériences 2 et 3**, nous avons étudié comment la présence d'une scène en vision périphérique peut influencer la catégorisation d'une scène en vision centrale. Nous avons pour objectif **(i)** de tester si la présence d'une scène en vision périphérique provoque une interférence sur les processus de catégorisation de la scène en vision centrale, **(ii)** d'étudier le rôle des caractéristiques physiques des scènes dans cet effet et **(iii)** de mettre en évidence les régions cérébrales associées à l'effet d'interférence périphérique.

2. Le terme d'interférence sémantique peut porter à confusion. En effet, il peut laisser penser que l'interférence se passe à un niveau sémantique ou implique des processus sémantiques. Ce paradigme expérimental sera donc désigné par le terme *paradigme d'interférence* dans le reste du document, et il en va de même pour l'effet mesuré (*effet d'interférence*, ou *effet d'interférence périphérique*).

- Dans les **Expériences 4 et 5**, nous avons de nouveau utilisé le paradigme d'interférence afin d'étudier comment l'information contextuelle d'une scène disponible en vision périphérique est susceptible d'influencer la reconnaissance d'un objet en vision centrale. Nous avons manipulé la congruence sémantique entre la scène et l'objet (e.g., un fauteuil était associé à une scène d'intérieur dans la condition congruente et à une scène d'extérieur dans la condition incongruente). Nous avons notamment pour objectif **(i)** de tester l'hypothèse d'une influence de l'extraction du *gist* en vision périphérique sur la reconnaissance d'objets, **(ii)** de tester l'hypothèse de mécanismes prédictifs dans cette influence, et **(iii)** de mieux comprendre la nature des influences, et notamment quel niveau de traitement de la scène (bas niveau *vs* sémantique) est nécessaire pour que celle-ci influence la reconnaissance des objets.

Dans les expériences du Chapitre 2 (mise à part l'Expérience 1) nous n'avons pas filtré les stimuli. Nous avons considéré que la présentation des stimuli en vision périphérique agirait comme un filtre passe-bas naturel. Par ailleurs, il est possible que la procédure de filtrage affecte certaines propriétés du signal, particulièrement lorsqu'il provient de la vision périphérique. Par exemple, il est possible que les mécanismes de *crowding* n'opèrent pas de la même manière sur une image filtrée passe-bas que sur une image intacte. Le fait de ne pas filtrer les scènes permet ainsi de mieux imiter le signal naturel et de préserver au maximum le fonctionnement des mécanismes propres à la vision périphérique.

4.5. Le glaucome comme modèle pathologique d'une perturbation de la vision périphérique

Le second volet empirique de ces travaux de thèse est présenté dans le Chapitre 3 (**Expériences 6 et 7**). Les expériences que nous reportons dans ce chapitre ont été réalisées auprès de patients atteints de glaucome. Le glaucome est une maladie ophtalmologique qui détruit progressivement les cellules ganglionnaires de la rétine et provoque des altérations du champ visuel, majoritairement au niveau du champ visuel périphérique. Cette maladie représente ainsi un bon modèle pathologique pour étudier les influences de la vision périphérique sur la reconnaissance en vision centrale. Comme la maladie est causée par une destruction des cellules ganglionnaires, la plupart des études sur le glaucome se concentre sur des aspects très bas niveau du traitement visuel (e.g., détection de points lumineux dans le champ visuel), et les phénomènes de plus haut niveau comme la reconnaissance visuelle ont été à ce jour assez peu étudiés. Nous pensons pourtant que l'atteinte des cellules ganglionnaires dans le glaucome, causant une perte partielle et principalement périphérique du champ visuel, pourrait impacter la fonction visuelle dans son ensemble. Plus spécifiquement, notre hypothèse est que les mécanismes de reconnaissance de stimuli perçus en vision centrale devraient être perturbés par la perte de la vision périphérique.

Le glaucome survient la plupart du temps chez le sujet âgé. En raison de l'âge des participants, de leur maladie et de leur manque de pratique dans les expériences de psychologie, il est relativement difficile de leur demander des tâches similaires à celles demandées à des participants jeunes sans déficits visuels (qui de surcroît sont souvent des étudiants de psychologie habitués aux conditions expérimentales). Aussi, nous avons pu constater avec quelques expériences pilotes qu'il était difficile, par exemple, de présenter des stimuli en vision périphérique, qui mettent les patients glaucomateux dans une situation d'échec et conduisent parfois à l'abandon. Nous avons alors favorisé les paradigmes expérimentaux les plus simples possibles. Comme les participants sains, les patients devaient catégoriser des stimuli visuels. Cependant, plutôt que de demander une catégorisation à deux choix forcés (ou 2AFC), nous leur avons demandé de donner leur réponse *via* une procédure go/no-go (i.e. ils devaient appuyer sur un unique bouton de réponse lorsqu'ils reconnaissaient le stimulus comme appartenant à une catégorie cible). Par ailleurs, nous avons présenté les stimuli seulement dans la vision centrale des participants. Nous nous attendions à ce que le modèle FP-péri soit perturbé chez les patients atteints de glaucome. Plus précisément, notre hypothèse est que le système de gestion des prédictions, recevant une information de moindre quantité et/ou qualité en provenance de la rétine périphérique, se modifie peu à peu et devient moins efficace. Ainsi, même l'information extraite en vision centrale ne pourrait pas être traitée avec la même efficacité (Figure 1.21). Selon cette hypothèse et sur la base de résultats de la littérature, il devrait être possible de mesurer des déficits - possiblement subtils - de reconnaissance visuelle dans la vision centrale des patients glaucomateux. Les expériences que nous avons menées auprès de patients glaucomateux dans l'objectif de tester cette hypothèse sont les suivantes :

- Dans l'**Expérience 6**, nous avons présenté aux patients et à des participants contrôles des stimuli très faiblement contrastés dans leur vision centrale. Cette manipulation était motivée par des résultats récents indiquant que les patients glaucomateux peuvent montrer des difficultés à reconnaître des objets dont le contraste est diminué (Lenoble, Lek, & McKendrick 2016), mais également par le fait que ces patients se plaignent fréquemment d'avoir des difficultés pour reconnaître leur environnement dans des situations de luminosité extrêmes (e.g., la nuit, ou au contraire en cas de trop forte lumière). Le faible contraste de nos stimuli permettait ainsi d'imiter une situation d'obscurité. L'objectif de l'expérience était de comparer les performances à une tâche de détection des images (impliquant des processus de bas niveau) aux performances à une tâche de catégorisation (impliquant les processus de plus haut niveau permettant la reconnaissance visuelle).
- L'**Expérience 7** était une étude exploratoire en IRMf visant à explorer comment une désorganisation du modèle FP-péri pourrait se manifester au niveau cérébral. Nous avons adapté le protocole de l'Expérience 6 à l'IRMf afin d'étudier de manière exploratoire les changements fonctionnels cérébraux chez trois patients glaucomateux et trois participants contrôles appariés en âge lors de la

catégorisation en vision centrale de scènes visuelles faiblement contrastées.

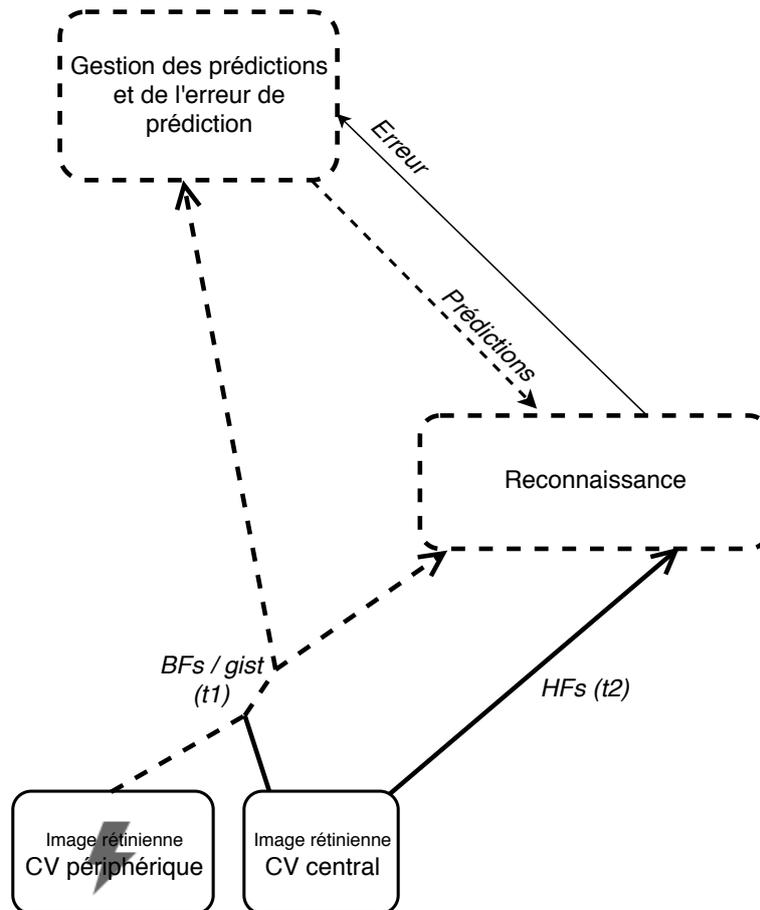


Figure 1.21: Modèle FP-peri dans le glaucome. L'éclair représente l'atteinte rétinienne glaucomateuse, qui entraînerait une perturbation du système de reconnaissance visuelle (les éléments perturbés du modèle sont représentés en pointillés). CV : champ visuel.

Chapitre 2

Contribution empirique :

*Influence de la vision périphérique sur les
traitements réalisés en vision centrale
chez le sujet sain*

1. Préambule pour le traitement des données et l'analyse statistique

Nous avons suivi la même démarche, détaillée ci-après, pour l'analyse des données comportementales des cinq expériences présentées dans ce Chapitre.

Pour les cinq expériences nous avons utilisé deux mesures : les temps de réponses correctes et la proportion de réponses correctes, que nous avons analysées à l'aide de modèles à effets mixtes. Les modèles mixtes permettent de modéliser non seulement les effets fixes, comme dans une analyse plus classique (e.g., modèle linéaire ou analyse de la variance), mais aussi les effets aléatoires. Les effets fixes sont ceux de nos variables indépendantes d'intérêt. Les effets aléatoires sont les effets liés à la variabilité des participants, ou éventuellement à la variabilité des stimuli. Ici, nous avons pris en compte seulement les effets aléatoires liés aux participants et pas ceux liés aux stimuli, étant donné que ces derniers n'ont pas permis un meilleur ajustement des modèles dans la plupart des cas. Les effets aléatoires comprennent l'intercept (ou ordonnée à l'origine) de chaque participant, et la variabilité des réponses de chaque participant. L'intérêt de prendre en compte les effets aléatoires est que l'estimation des paramètres du modèle au niveau du groupe est plus stable et moins impactée par des participants peu fiables (e.g., un participant qui a une grande variabilité à travers les essais, ou un autre qui a des temps de réponse moyens considérablement plus longs que les autres participants). Par ailleurs, il faut souligner ici que certaines des expériences de ce chapitre ont été publiées dans des revues scientifiques (Expérience 1 dans la revue *Visual Cognition* - [Kauffmann et al. 2017](#), et Expériences 4 et 5 dans la revue *Journal of Vision* - [Roux-Sibilon et al. 2019](#)). Lorsque nous avons analysé les données de ces expériences pour leur publication, nous avons alors utilisé des analyses de la variance à mesures répétées, avec une approche d'analyse exhaustive du plan factoriel. Nous entendons par là que nous avons réalisé les tests des effets principaux des facteurs intra-sujets et inter-sujets, leurs effets d'interaction, et, lorsque les effets d'interaction étaient significatifs, les tests des effets simples. Cette approche a l'avantage de permettre une large liberté d'analyse au sein de plans factoriels complexes. Elle permet, par exemple, de tester les effets prédits mais également de tester des effets non prédits, dans une démarche exploratoire. Cependant, elle a comme inconvénient, en envisageant des interactions impliquant le nombre maximum de facteurs du plan expérimental (e.g., dans l'Expérience 1, quatre facteurs), de rendre la lecture et la compréhension des résultats parfois complexe. Par ailleurs, cette approche est peu parcimonieuse et multiplie le nombre de tests effectués (pouvant potentiellement augmenter le risque d'erreur de type 1 au delà du seuil de 5%). Dans ce manuscrit de thèse, nous choisissons donc de ré-analyser ces données en testant statistiquement les seuls effets prédits. De plus, nous utilisons des modèles à effets mixtes afin de pouvoir prendre en compte les effets aléatoires liés aux participants, comme nous l'avons indiqué précédemment. Globalement, pour les expériences publiées (Expériences 1, 4 et 5), ces

deux types d'analyse permettent de tirer les mêmes conclusions.

1.1. Analyse de la proportion de réponses correctes

Pour analyser la proportion de réponses correctes, nous avons utilisé des modèles de régression logistique mixtes. Les modèles logistiques sont utilisés pour les variables dépendantes binaires (ici, 1 pour une réponse correcte, et 0 pour une réponse incorrecte ou une absence de réponse ; [Gelman & Hill 2006](#); [Jaeger 2008](#)).

Dans ces analyses, nous porterons notre attention sur deux choses : (1) le test de significativité, qui nous permet de conclure sur la présence ou non de chacun des effets, et (2) le paramètre estimé bêta (β , la pente) des modèles, i.e. l'estimation par le modèle de l'effet dans la population, qui reflète la magnitude de l'effet. Pour chaque modèle (chaque effet testé), nous reporterons donc le paramètre β et le test de significativité pour les effets fixes seulement (les effets aléatoires n'ayant que peu d'intérêt en tant que tels ici). La significativité des effets a été testée avec un test de Wald et décidée à un seuil alpha de 0.05. Ces analyses ont été réalisées avec le package *lme4* ([Bates, Maechler, Bolker, Walker, et al. 2014](#)) dans l'environnement R.

Le paramètre β d'une régression logistique indique de combien augmente le logit des chances (*odds*) de faire une réponse correcte plutôt qu'incorrecte dans une des conditions par rapport à l'autre. Cette valeur n'est pas interprétable telle quelle, et il est d'usage de calculer un rapport des chances (*odds ratio*) à partir du paramètre β . L'interprétation des *odds ratio* (OR) est assez difficile puisqu'il s'agit d'un ratio de ratios (voir la [Figure 2.1](#) pour un exemple représentatif des expériences du Chapitre 2). Cependant, on peut globalement garder en tête, pour la lecture des résultats, qu'un $OR > 1$ signifie que les chances de faire une réponse correcte sont plus importantes dans la condition 1 que dans la condition 2, qu'un OR proche de 1 signifie que les chances de faire une réponse correcte sont similaires dans les deux conditions, et qu'un $OR < 1$ signifie que les chances de faire une réponse correcte sont plus importantes dans la condition 2 que dans la condition 1. Les OR pouvant prendre des valeurs de *-infini* à *+infini*, nous pourrions avoir une idée de la taille des effets en fonction de la distance à 1 du OR.

1.2. Analyse des temps de réponses correctes

Pour analyser les temps de réponses correctes, nous avons utilisé des modèles mixtes linéaires. Avant de modéliser les temps de réponses correctes, ceux-ci étaient d'abord écrémés pour chaque participant et chaque condition expérimentale afin de limiter l'impact des valeurs extrêmes. Nous avons pour cela retiré les temps de réponses supérieurs ou inférieurs à la moyenne ± 2.5 écart-types dans chaque condition expérimentale (i.e. chaque croisement orthogonal des facteurs d'intérêts). Nous avons ensuite transformé les temps de réponses correctes par une fonction logarithmique afin que les résidus des modèles statistiques puissent être distribués normalement. Ici encore, nous porterons notre attention sur (1) le test de significativité et (2) le paramètre estimé β pour les

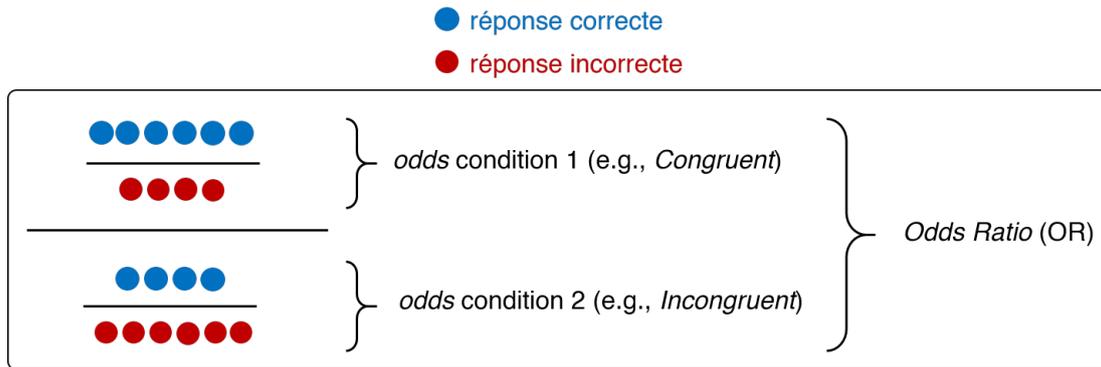


Figure 2.1: Représentation schématisque du *odds ratio*. Dans cet exemple, les *odds* de faire une réponse correcte pour la condition 1 (Congruent) sont de $6/4 = 1.5$, et les *odds* de faire une réponse correcte pour la condition 2 (Incongruent) sont de $4/6 = 0.66$. Le *odds ratio* (OR) correspond ainsi au rapport entre ces deux valeurs : $1.5/0.66 = 2.24$. C'est donc ce rapport qui renseigne sur le sens et surtout la taille de l'effet. Pour prendre un autre exemple, si les *odds* de faire une réponse correcte pour la condition 1 (Congruent) sont de $7/3 = 2.33$, et les *odds* de faire une réponse correcte pour la condition 2 (Incongruent) sont de $5/5 = 0.50$, l'OR sera de $2.33/0.50 = 4.66$, montrant un effet de plus grande taille que dans le premier exemple.

effets fixes. Du fait de la transformation logarithmique des données, le paramètre β est exprimé en unités log. Sur cette base, nous avons calculé un pourcentage de changement d'une condition expérimentale à l'autre. Cette valeur, notée *%change*, sera reportée pour rendre compte de la taille des effets. Par exemple, si le *%change* calculé sur la base du paramètre β d'un modèle est de 5%, on peut estimer que (dans la population) les temps de réponse moyens augmentent de 5% dans la condition 2 par rapport à la condition 1. La significativité des effets a été testée en ajustant les modèles au maximum de vraisemblance restreint en déterminant les degrés de liberté par approximation Satterthwaite, et décidée à un seuil alpha de 0.05. Ces analyses sont réalisées avec les packages *lme4* (Bates et al. 2014) et *lmerTest* (Kuznetsova, Brockhoff, & Christensen 2017) dans l'environnement R. À noter que nous avons modélisé les temps de réponses correctes transformés ($\log(\text{TR})$), mais que les graphiques représentent les temps de réponses correctes bruts moyens (TRm, non transformés) en millisecondes.

2. Expérience 1 - Étude de l'intégration des basses et hautes fréquences spatiales dans des scènes hybrides

N.B. Dans cette partie expérimentale, nous reportons les données d'une expérience que nous avons publiée dans la revue *Visual Cognition* (Kauffmann et al. 2017; Annexe B). Dans cet article, nous avons testé nos hypothèses à l'aide d'une analyse de la variance à mesures répétées. Comme indiqué en préambule, nous avons ré-analysé ces données en utilisant des modèles à effets mixtes à l'occasion de la rédaction de ce manuscrit de thèse (cf. 1, Préambule).

2.1. Objectifs et hypothèses

Cette expérience avait pour but d'étudier comment le traitement des informations contenues dans les BFs influence le traitement des HFes lors de la catégorisation de scènes, mais aussi de tester comment le traitement des HFes peut en retour influencer le traitement des BFs. Nous avons utilisé le paradigme d'interférence. Nous avons présenté aux participants des scènes hybrides construites en combinant les BFs d'une image de scène avec les HFes d'une autre image de scène (Kauffmann et al. 2015c; Mu & Li 2013; Schyns & Oliva 1994). La moitié des participants devait catégoriser la scène en HFes de l'hybride, tout en ignorant la scène en BFs. L'autre moitié des participants devait catégoriser la scène en BFs, et ignorer la scène en HFes. Les participants devaient catégoriser la scène cible de l'hybride comme appartenant soit à la catégorie 'naturelle' (e.g., montagnes, plages), soit à la catégorie 'artificielle' (e.g., villes, autoroutes). Les catégories que les participants devaient former étaient donc sur-ordonnées, la catégorisation s'appuyant sur les propriétés globales des scènes. Dans la moitié des essais, les deux scènes composant chaque hybride appartenaient à la même catégorie sémantique (condition sémantiquement congruente; e.g., les BFs d'une scène de plage et les HFes d'une scène de montagne), tandis que dans l'autre moitié des essais elles appartenaient à des catégories sémantiques différentes (condition sémantiquement incongruente; e.g., les BFs d'une scène de plage et les HFes d'une scène d'autoroute; cf. Figure 2.2). En comparant ces deux conditions, nous pouvons donc examiner si le contenu en fréquences spatiales à "ignorer" est tout de même traité, et comment son traitement influence celui du contenu en fréquences spatiales cible. Cette tâche avec contrainte attentionnelle sur un contenu en fréquence spatiale cible était en fait réalisée lors d'une seconde session expérimentale (Session 2). Lors d'une première session expérimentale (Session 1), les participants devaient catégoriser les scènes filtrées en BFs et en HFes qui composaient les hybrides présentés cette fois-ci séparément. Cette session expérimentale avait pour but de nous assurer que la catégorisation des scènes BFs et HFes était possible, et de comparer leur traitement indépendamment de leur intégration et des contraintes attentionnelles.

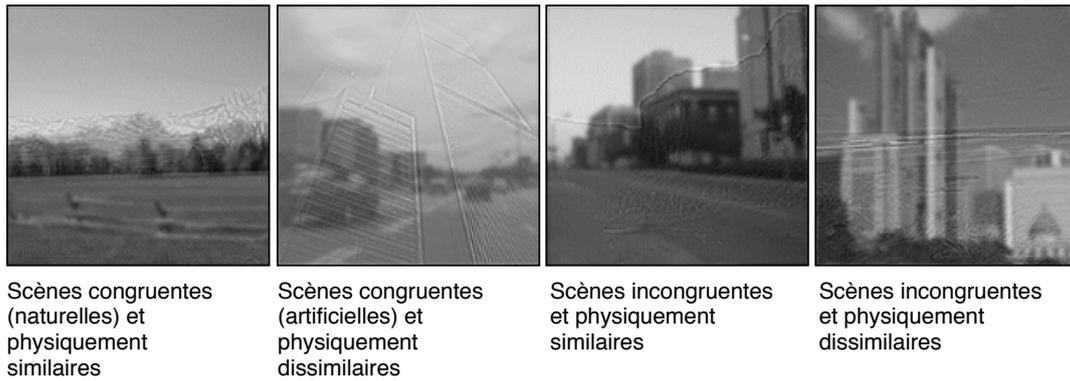


Figure 2.2: Types d'images hybrides utilisées dans l'Expérience 1. Les hybrides étaient formés par la superposition des basses fréquences spatiales d'une scène et des hautes fréquences spatiales d'une autre scène. Les deux scènes étaient sémantiquement congruentes (i.e. deux scènes artificielles ou deux scènes naturelles) ou sémantiquement incongruentes (une scène artificielle et une scène naturelle). De plus, les hybrides congruents et incongruents pouvaient être constitués de deux scènes physiquement similaires ou de deux scènes physiquement dissimilaires (cf. partie 2.1.3).

Afin de faciliter la lecture des résultats de la Session 2 (scènes hybrides), nous appellerons **effet d'interférence** l'effet de la congruence sémantique entre les deux scènes qui composent une image hybride. L'effet d'interférence se caractérise par des temps de réponses correctes plus courts et/ou une plus grande proportion de réponses correctes pour la condition congruente que pour la condition incongruente. Nous analyserons nos données en fonction de trois axes d'analyse principaux, qui sont abordées ci-après, et des hypothèses qui y sont associées.

2.1.1. Précédence temporelle du traitement des BFs lors de la catégorisation de scènes, et rôle du temps de traitement

Dans le cadre du modèle FP, nous avons fait l'hypothèse que la catégorisation d'une scène en BFs (qu'elle soit présentée individuellement dans la Session 1 ou superposée à une scène en HFfs dans la Session 2) devrait être plus précise et plus rapide que la catégorisation d'une scène en HFfs. Nous avons donc testé l'effet principal de la fréquence spatiale (BFs *vs* HFfs) de la scène à catégoriser en Session 1, et l'effet principal de la fréquence spatiale cible de l'image hybride (cible-BFs *vs* cible-HFfs) en Session 2.

Par ailleurs, nous avons examiné dans quelle mesure ces effets dépendent du temps de traitement des stimuli. Ceux-ci (scènes filtrées dans la Session 1 ou scènes hybrides dans la Session 2) étaient présentés pendant 30 ms ou pendant 100 ms, afin d'examiner comment le poids relatif des BFfs et des HFfs peut varier au cours du temps. Nous avons donc testé l'effet d'interaction entre la fréquence spatiale et le temps de présentation dans la Session 1 et l'effet d'interaction entre la fréquence spatiale cible et le temps de présentation dans la Session 2.

2.1.2. Influences mutuelles des BFs et des HFes lors de la catégorisation des scènes hybrides

Selon le modèle FP, mais aussi dans la lignée des résultats de [Mu et Li \(2013\)](#), puis de [Kauffmann et al. \(2015c\)](#), nous nous attendions d'abord à un effet d'interférence dans la condition cible-HFs dans la Session 2, c'est-à-dire à une interférence des BFs sur la catégorisation de la scène en HFes de l'image hybride. Par ailleurs, cette expérience avait également pour objectif de tester l'existence d'une éventuelle interférence des HFes sur le traitement des BFs, autrement dit comment l'information en HFes peut en retour influencer la manière dont les BFs sont traitées. Comme nous l'avons vu, il existe également des arguments en faveur de l'hypothèse d'un traitement prédominant des HFes même lorsque les temps de présentation des stimuli visuels sont très courts (e.g., [Özgen et al. 2005](#); [Rotshtein et al. 2010](#); [Schyns & Oliva 1999](#), cf. Chapitre 1, partie 2.2.3). Le poids des HFes pourrait donc malgré tout être important lors de la catégorisation de scènes et influencer en retour le traitement des BFs. Toutefois, les études qui ont mis en évidence une utilisation flexible des fréquences spatiales se sont concentrées sur l'utilisation préférentielle d'une bande de fréquence spatiale par rapport à une autre, ou sur leur diagnosticité respective pour une tâche ou une condition donnée. Peu d'études se sont intéressées à l'impact d'un traitement rapide des HFes sur le traitement des BFs. Dans leur article, [Mu et Li \(2013\)](#) ont rapporté succinctement les résultats d'une expérience contrôlée dans laquelle les participants devaient catégoriser la scène en BFs de l'image hybride, tout en ignorant la scène en HFes (i.e. le schéma inverse de leur expérience principale où les participants devaient catégoriser les HFes tout en ignorant les BFs). Ils ont observé une interférence, indiquant que l'information en HFes est également susceptible d'influencer la catégorisation des scènes en BFs. Mais les performances de catégorisation des scènes en BFs étaient au niveau du hasard dans la condition d'incongruence sémantique entre les deux scènes de l'image hybride, indiquant que les participants n'étaient en fait pas capables de réaliser la tâche dans cette condition. Par ailleurs, les auteurs n'ont pas comparé l'interférence des HFes sur les BFs à l'interférence des BFs sur les HFes. Il nous semblait donc important de comparer directement ces deux conditions attentionnelles (cible-BFs *vs* cible-HFs). Nous avons donc testé l'effet d'interaction entre la congruence et la fréquence spatiale cible. Par ailleurs, nous avons examiné comment cet effet varie en fonction du temps de traitement des hybrides (double effet d'interaction entre la congruence, la fréquence spatiale cible et le temps de présentation). Nous avons fait l'hypothèse suivante : pour un temps de présentation court (30 ms), l'interférence des BFs sur la catégorisation des HFes devrait être plus forte que l'interférence des HFes sur la catégorisation des BFs. Cet effet d'interaction devrait être réduit pour un temps de présentation plus long (100 ms), voire même inversé de telle manière à ce que l'effet d'interférence des HFes soit plus fort que celui des BFs.

2.1.3. Nature des influences : Effet de la similarité physique

Nous nous sommes aussi intéressés à la nature de l'influence prédictive, en manipulant la similarité physique entre les scènes BFs et HFs de l'image hybride. La similarité physique portait à la fois sur la configuration spatiale des scènes et sur le spectre d'amplitude. En d'autres termes, deux scènes physiquement similaires se caractérisent **(i)** par une correspondance entre les valeurs d'intensité de leurs pixels et **(ii)** par une similarité entre leurs distributions de luminance sur les différentes fréquences spatiales et orientations. Deux scènes sémantiquement congruentes pouvaient ainsi être soit physiquement similaires (configuration spatiale et spectre d'amplitude similaires), soit physiquement dissimilaires. De même, deux scènes sémantiquement incongruentes pouvaient être soit physiquement similaires, soit physiquement dissimilaires (Figure 2.2). Nous pouvons ainsi examiner comment la similarité physique entre les deux scènes d'une image hybride module l'effet d'interférence. Au niveau théorique, le système de gestion des prédictions du modèle FP pourrait se servir des propriétés physiques de la scène distractive en BFs pour former des prédictions sur la catégorie de la scène ainsi que sur la présence d'éléments physiques (e.g., une orientation dominante) à des endroits précis de la scène (e.g., en bas de l'image). Si le système de prédictions utilise effectivement ces informations physiques, l'intégration des prédictions et des informations ascendantes dans le système de reconnaissance devrait être plus importante lorsque leurs propriétés physiques sont semblables.

Sur la base de cette hypothèse et des résultats de [Mu et Li \(2013\)](#), nous nous attendions à un effet d'interférence plus grand lorsque les scènes filtrées composant les hybrides sont physiquement similaires que dissimilaires. En d'autres termes, nous nous attendions à ce que la présence de la scène distractive influence encore plus la catégorisation de la scène cible lorsque ses propriétés physiques sont semblables. Par parcimonie et parce que nous n'avions pas a priori d'hypothèse forte d'interaction avec la fréquence spatiale cible ou le temps de présentation, nous avons testé si l'effet d'interférence varie en fonction de la similarité physique d'abord globalement, puis indépendamment pour chaque condition croisée de fréquence spatiale cible et de temps de présentation.

2.2. Méthode

2.2.1. Participants

Au total, 52 étudiants de psychologie ont participé à l'Expérience 1 (38 femmes, âge moyen \pm écart-type : 22 ± 3 ans). Tous étaient droitiers et avaient une vision normale ou corrigée à la normale et pas d'antécédents de maladies neurologiques ou psychiatriques. Avant de participer à l'expérience, chaque participant a donné son consentement libre et éclairé en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine. À la fin de l'expérience, chaque participant recevait un bon d'expérience de 0.5 point à faire valoir lors des examens universitaires.

2.2.2. Stimuli et procédure

Les images utilisées pour la construction des stimuli hybrides étaient des photographies couleur de scènes visuelles (résolution 256×256 pixels) provenant de la base d'images *Labelme* du *Computational Visual Cognition Laboratory* (Oliva & Torralba 2001). La manipulation des images a été réalisée dans l'environnement MATLAB (Mathworks Inc., Sherborn, MA, USA). Nous avons tout d'abord classifié les 2687 scènes de la base d'images en deux catégories distinctes : les scènes 'naturelles' (e.g., plages, champs) et les scènes 'artificielles' (e.g., villes, autoroutes). Nous avons ensuite évalué la similarité physique des scènes sur la base de (i) la corrélation 'IP' entre les valeurs d'intensité des pixels et (ii) la corrélation 'SA' entre les valeurs d'intensité des pixels des spectres d'amplitude des scènes (pour une procédure similaire, voir Kauffmann et al. 2015c). Nous avons considéré que deux scènes étaient physiquement similaires si ces deux coefficients IP et SA étaient supérieurs à 0.6 (seulement 1.59% des paires d'images), et nous avons considéré que deux scènes étaient physiquement dissimilaires si le coefficient de corrélation IP était inférieur à 0.01 et le coefficient de corrélation SA était inférieur à 0.6 (seulement 1.37% des images). Parmi les paires qui avaient des propriétés physiques similaires, nous avons sélectionné 40 paires de scènes sémantiquement congruentes (20 paires de scènes 'naturelles' et 20 paires de scènes 'artificielles'; IP : 0.75 ± 0.08 ; SA : 0.67 ± 0.03) et 40 paires de scènes sémantiquement incongruentes (20 paires de scènes 'naturelles' et 20 paires de scènes 'artificielles'; IP : 0.73 ± 0.06 ; SA : 0.66 ± 0.03). Par ailleurs, les scènes étaient sélectionnées pour avoir des valeurs de corrélations IP et SA comparables entre les paires sémantiquement congruentes et incongruentes. De la même manière, parmi les paires qui avaient des propriétés physiques dissimilaires, nous avons sélectionné 40 paires de scènes sémantiquement congruentes (20 paires de chaque catégorie; IP : 0.01 ± 0.00 ; SA : 0.43 ± 0.06) et 40 paires de scènes sémantiquement incongruentes (IP : 0.01 ± 0.00 ; SA : 0.44 ± 0.05). Ici aussi, les scènes étaient sélectionnées pour avoir des valeurs de corrélations IP et SA comparables entre les paires sémantiquement congruentes et incongruentes. Au total, nous avons donc formé 160 paires de scènes.

Nous avons ensuite converti les images en 256 niveaux de gris et nous les avons filtrées afin de conserver l'information en BF's ou en HF's. Pour appliquer la procédure de filtrage, nous avons fixé la taille angulaire des images à $6^\circ \times 6^\circ$ d'angle visuel. Les images filtrées étaient obtenues en multipliant la transformée de Fourier des images originales avec des filtres gaussiens. La largeur des filtres gaussiens (i.e. leurs écart-types) dépendait de la fréquence de coupure pour une atténuation du signal de 3dB (soit l'atténuation habituellement utilisée en traitement du signal). L'image en BF's ne conservait que les fréquences spatiales inférieures ou égales à 24 cpi (4 cpd). L'image en HF's ne conservait que les fréquences spatiales supérieures ou égales à 36 cpi (6 cpd). Nous avons ensuite égalisé la luminance moyenne de chaque image filtrée à 0.5 pour des valeurs d'intensité des pixels allant de 0 à 1 (c'est-à-dire une luminance moyenne de 128 sur une échelle de 256 niveaux de gris). Avec cette procédure, les pixels de certaines images avaient

des valeurs supérieures à 1 ou inférieures à 0. Nous avons réajusté ces valeurs en les ramenant à 1 ou à 0, respectivement. Cette correction était minimale et la valeur moyenne de luminance des scènes filtrées était tout de même de 0.5 après correction. À noter qu'à l'issue du filtrage, le contraste de luminance des scènes en BFs était plus élevé que celui des scènes en HFs (contraste *Root-Mean-Square* ou RMS moyen \pm écart-type pour les scènes BFs 'naturelles' : 0.24 ± 0.06 ; scènes BFs 'artificielles' : 0.24 ± 0.05 ; scènes HFs 'naturelles' : 0.06 ± 0.02 ; scènes HFs 'artificielles' : 0.07 ± 0.02).

Les stimuli hybrides ont été créés en couplant une scène en BFs avec une autre scène en HFs, l'association de deux scènes dépendant des conditions de congruence sémantique et de similarité physique. Les deux scènes d'une image hybride pouvaient donc être (1) sémantiquement congruentes et physiquement similaires (congruent-PHYSIM), (2) sémantiquement congruentes et physiquement dissimilaires (congruent-PHYDIS), (3) sémantiquement incongruentes et physiquement similaires (incongruent-PHYSIM) ou (4) sémantiquement incongruentes et physiquement dissimilaires (incongruent-PHYDIS). Comme nous l'avons mentionné plus haut, les 160 scènes filtrées qui composaient les images hybrides ont été aussi utilisées dans une première session expérimentale (Session 1) où elles étaient présentées séparément (soit dans leur version BFs, soit dans leur version HFs, voir Figure 2.3).

L'expérience était programmée avec le logiciel de conception d'expérience et de recueil de données E-prime 2 (E-prime Psychology Software Tools Inc., Pittsburgh, USA). Les stimuli apparaissaient sur un écran d'ordinateur cathodique 17 pouces d'une résolution de 1025×768 pixels, avec un taux de rafraîchissement de 80 Hz. Les stimuli étaient affichés sur un fond gris uniforme de 117 sur une échelle de 256 niveaux de gris. Les participants étaient assis dans l'obscurité à une distance de 73 cm de l'écran. À cette distance, les stimuli respectaient la taille de $6^\circ \times 6^\circ$ d'angle visuel. Pour maintenir une position centrée et la distance à l'écran, la tête des participants reposait sur une mentonnière positionnée dans l'axe sagittal médian.

Chaque participant passait deux sessions expérimentales. Dans la Session 1, les stimuli étaient les scènes filtrées en BFs et HFs. Les participants étaient répartis en deux groupes ($N = 26$ dans chaque groupe). Les scènes étaient présentées pendant 30 ms pour le premier groupe, et pendant 100 ms pour le second. Les scènes BFs et HFs étaient présentées de manière aléatoire. Les participants devaient catégoriser chaque scène filtrée comme appartenant soit à la catégorie 'naturelle', soit à la catégorie 'artificielle'. Dans la Session 2, les stimuli étaient les images hybrides. Pour chaque groupe affecté à un temps de présentation lors de la Session 1, la moitié des participants ($N = 13$) avait pour consigne d'ignorer la scène en BFs et de catégoriser la scène en HFs (cible-HFs; les participants étaient ainsi dans des conditions d'expérimentation similaires à celles des études précédentes; [Kauffmann et al. 2015c](#); [Mu & Li 2013](#)), tandis que l'autre moitié recevait la consigne inverse (cible-BFs).

Dans les deux sessions expérimentales, chaque essai se déroulait comme suit : le participant voyait d'abord un point de fixation central pendant 500 ms, puis le stimulus

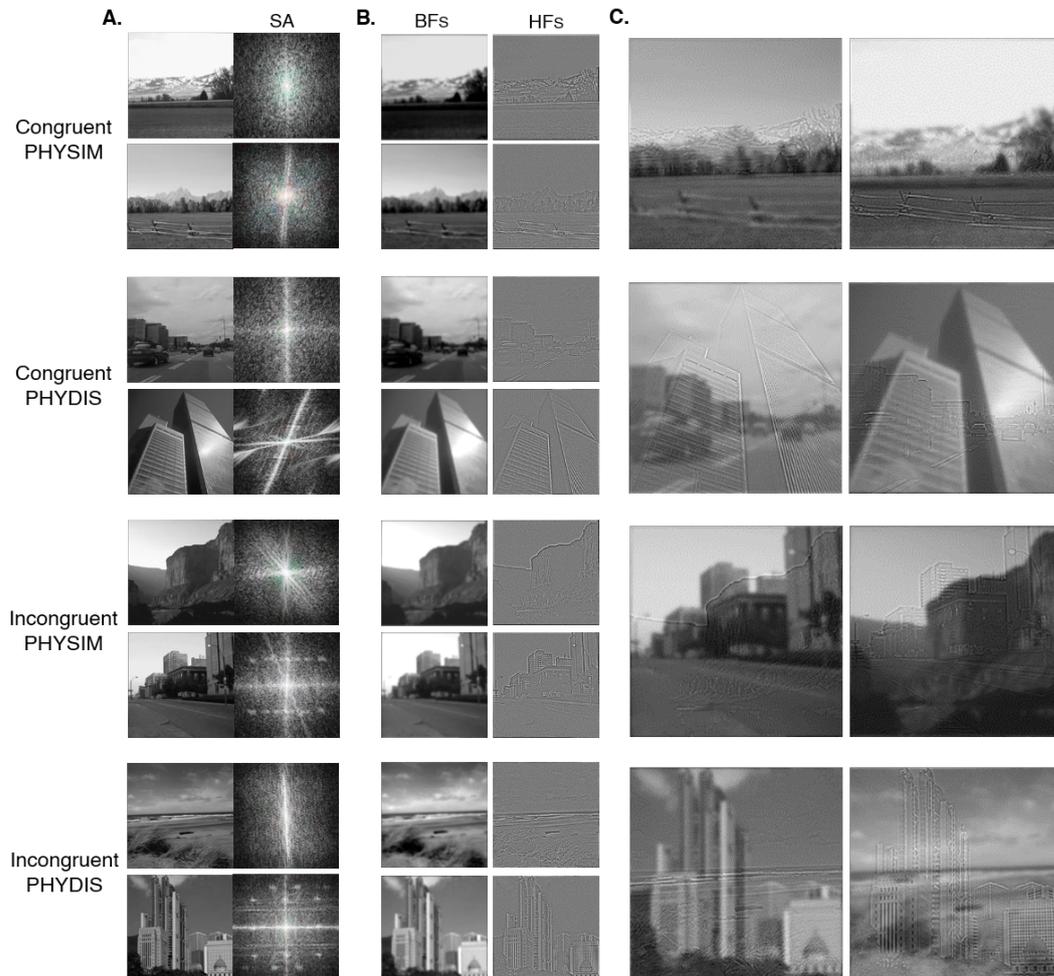


Figure 2.3: **A.** Exemples de scènes utilisées pour construire les stimuli de l'Expérience 1, avec leurs spectres d'amplitude (SA). Sur les spectres d'amplitudes, les BFs sont proches du centre et les HF's sont à la périphérie. **B.** Stimuli de la Session 1 : scènes artificielles et naturelles filtrées en BFs et en HF's. **C.** Stimuli hybrides de la Session 2. Les scènes hybrides étaient construites en superposant les BFs d'une scène aux HF's d'une autre scène. Les deux scènes composant chaque hybride pouvaient être (1) Congruentes et physiquement similaires (PHYSIM), (2) Congruentes et physiquement dissimilaires (PHYDIS), (3) Incongruentes et physiquement similaires (PHYSIM), ou (4) Incongruentes et physiquement dissimilaires (PHYDIS). Un groupe de participants devait catégoriser le composant BFs de l'hybride (condition cible-BFs) et un autre groupe de participants devait catégoriser le composant HF's de l'hybride (condition cible-HFs).

(scène filtrée dans la Session 1 ou image hybride dans la Session 2) pendant 30 ou 100 ms suivant le groupe expérimental. Le stimulus était suivi d'un masque (bruit distribué en $1/f$) présenté pendant 30 ms afin d'empêcher la persistance rétinienne, puis d'un écran gris pendant 1000 ms (Figure 2.4). Les participants répondaient aussi rapidement et précisément que possible à l'aide des touches 1 et 2 d'un clavier d'ordinateur. La correspondance entre les catégories et les touches de réponse était contrebalancée entre les participants. À chaque essai, l'exactitude de la réponse et le temps de réponse étaient enregistrés. L'expérience durait environ 30 minutes. Avant de commencer chaque session, les participants réalisaient quelques essais d'entraînement avec des stimuli qui n'étaient pas utilisés dans l'expérience afin de se familiariser aux stimuli et à la tâche.

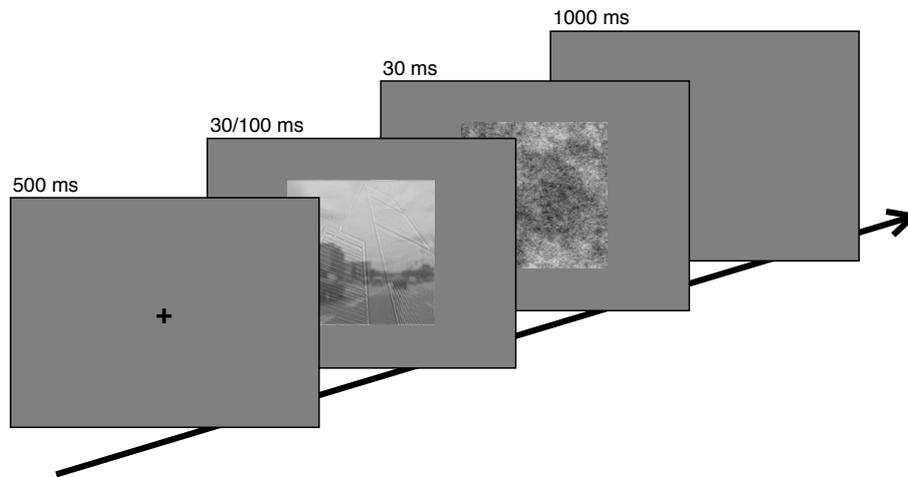


Figure 2.4: Procédure expérimentale de l'Expérience 1.

2.3. Résultats

2.3.1. Session 1 : Catégorisation des scènes filtrées

Pour l'analyse de la Session 1, les effets fixes des modèles mixtes ont été estimés pour la Fréquence spatiale (BFs, HF) et le Temps de présentation (30 ms, 100 ms), et leur interaction, et les effets aléatoires ont été pris en compte pour l'ordonnée à l'origine des participants.

Analyse de la proportion de réponses correctes

Les moyennes et écart-types du taux de réponses correctes pour chaque condition expérimentale sont reportés dans le Tableau 2.1. L'analyse montre un effet significatif de la Fréquence spatiale se traduisant par une plus grande proportion de réponses correctes lorsque les scènes sont filtrées en BFs que lorsqu'elles sont filtrées en HF (moyenne \pm écart-type : BFs = 0.92 ± 0.08 , HF = 0.88 ± 0.10 ; $\beta = -0.47$, OR = 1.60, $z = -4.45$, $p < .001$; Figure 2.5). La Fréquence spatiale n'interagit pas avec le Temps de présentation ($\beta = 0.35$, $z = 1.43$, $p = .154$).

2. Expérience 1 - Étude de l'intégration des basses et hautes fréquences spatiales dans des scènes hybrides

Tableau 2.1: Résultats de l'Expérience 1 (Session 1). Moyennes \pm écart-types de la proportion de réponses correctes (pRC), des temps de réponses correctes moyens en millisecondes (TRm) et des TR transformés (transformation logarithmique) moyens ($\log(\text{TR})_m$) en fonction du Temps de Présentation des scènes (30 ms, 100 ms) et de la Fréquence spatiale (BFs, HFfs)

	30 ms - BFs	30 ms - HFfs	100 ms - BFs	100 ms - HFfs
pRC	0.87 ± 0.08	0.80 ± 0.09	0.96 ± 0.04	0.95 ± 0.05
TRm (ms)	626 ± 83	667 ± 101	549 ± 49	576 ± 72
$\log(\text{TR})_m$	6.41 ± 0.13	6.47 ± 0.14	6.29 ± 0.09	6.33 ± 0.11

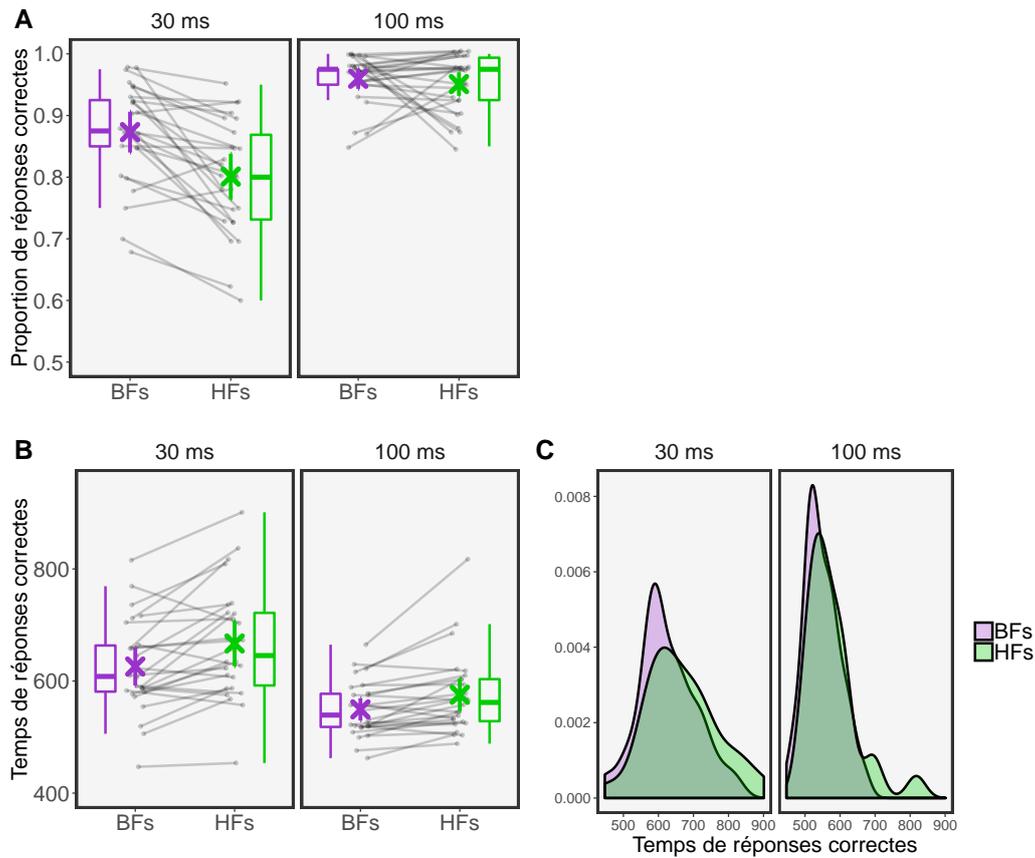


Figure 2.5: Résultats de l'Expérience 1 (Session 1). **A.** Proportion de réponses correctes et **B.** Temps de réponses correctes (ms) en fonction de la Fréquence spatiale (BFs ou HFfs) et du Temps de présentation des scènes (30 ms ou 100 ms). Les \times représentent les moyennes et les barres d'erreur qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet pour chaque participant. **C.** Distributions comparées des temps de réponses correctes pour les conditions BFs et HFfs.

Analyse des temps de réponses correctes

Les moyennes et écart-types des TRm et des $\log(\text{TR})_m$ pour chaque condition expérimentale sont reportés dans le Tableau 2.1. Suite à la procédure d'écrémage, 2.68% des temps de réponse ont été retirés des données utilisées pour les analyses statistiques. Le

modèle montre un effet significatif de la Fréquence spatiale se traduisant par des TRm plus courts pour catégoriser des scènes en BFs qu'en HFs (BFs = 588 ms \pm 78, HFs = 621 ms \pm 99; $\beta = 0.05$, %change = 5.00%, $t(3573) = 7.41$, $p < .001$; Figure 2.5). La Fréquence spatiale n'interagit pas avec le Temps de présentation ($\beta = -0.02$, $t(3572) = -1.22$, $p = .225$).

Effet du groupe (cible-BFs et cible-HFs)

Afin de s'assurer que les participants qui constitueraient les groupes cible-BFs et cible-HFs dans la Session 2 catégorisaient les scènes en HFs et en BFs de manière similaire entre les deux groupes, nous avons représenté les résultats de la Session 1 en fonction du groupe. Les patterns de réponse des deux groupes de participants (cible-BFs et cible-HFs) sont représentés dans la Figure 2.6. Au niveau des proportions de réponses correctes, les deux groupes se comportent de manière très similaire. Au niveau des temps de réponses correctes, on constate une plus grande variabilité des participants du groupe cible-HFs par rapport au groupe cible-BFs, lorsque les scènes sont présentées pendant 100 ms. Cependant, quel que soit le groupe, les effets semblent, pour les deux groupes, aller dans le sens d'un avantage des BFs par rapport aux HFs.

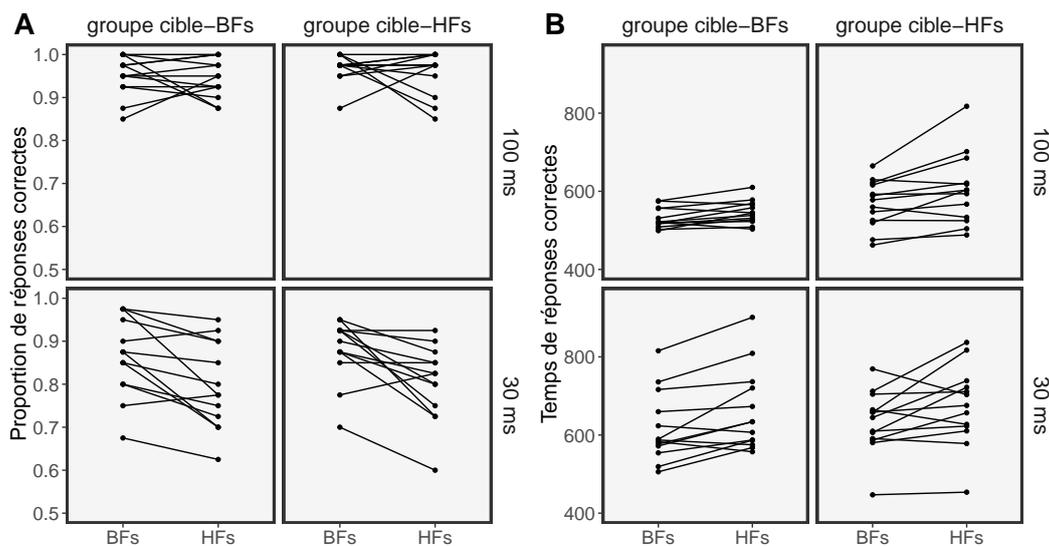


Figure 2.6: Comportement des groupes cible-BFs et cible-HFs de la Session 2 pendant la Session 1. **A.** Proportion de réponses correctes et **B.** Temps de réponses correctes (ms) en fonction de la Fréquence spatiale (BFs - HF), du Temps de présentation des scènes (30 ms - 100 ms), et du groupe de participants (cible-BFs - cible-HFs). Les droites correspondent à la pente de l'effet pour chaque participant.

2.3.2. Session 2 : Catégorisation des scènes hybrides

Pour l'analyse de la Session 2, les effets fixes des modèles mixtes ont été estimés pour le Temps de présentation (30 ms, 100 ms), la Fréquence spatiale cible (cible-BFs, cible-HFs), la Congruence (congruent, incongruent) et la Similarité physique (PHYSIM,

2. Expérience 1 - Étude de l'intégration des basses et hautes fréquences spatiales dans des scènes hybrides

PHYDIS) entre les scènes, et leurs interactions, et les effets aléatoires ont été pris en compte pour l'ordonnée à l'origine des participants.

Tableau 2.2: Résultats de l'Expérience 1 (Session 2). Moyennes \pm écart-types de la proportion de réponses correctes (pRC), des temps de réponses correctes moyens en millisecondes (TRm) et des TR transformés (transformation logarithmique) moyens (log(TR)m) en fonction du Temps de présentation des scènes (30 ms ; 100 ms), de la Fréquence spatiale cible (cible-HFs ; cible-HFs), de la Similarité physique entre les scènes (PHYSIM : physiquement similaire ; PHYDIS : physiquement dissimilaire) et de la Congruence sémantique (congruent ; incongruent)

30 ms								
	cible-HFs PHYSIM Congruent	cible-HFs PHYSIM Incongruent	cible-HFs PHYDIS Congruent	cible-HFs PHYDIS Incongruent	cible-BFs PHYSIM Congruent	cible-BFs PHYSIM Incongruent	cible-BFs PHYDIS Congruent	cible-BFs PHYDIS Incongruent
p(RC)	0.71 \pm 0.17	0.61 \pm 0.10	0.72 \pm 0.18	0.59 \pm 0.15	0.81 \pm 0.01	0.55 \pm 0.05	0.82 \pm 0.10	0.60 \pm 0.09
TRm (ms)	821 \pm 182	813 \pm 142	808 \pm 175	831 \pm 163	630 \pm 92	705 \pm 101	632 \pm 80	687 \pm 155
log(TR)m	6.66 \pm 0.21	6.65 \pm 0.18	6.63 \pm 0.21	6.67 \pm 0.20	6.41 \pm 0.15	6.52 \pm 0.15	6.41 \pm 0.15	6.48 \pm 0.23
100 ms								
	cible-HFs PHYSIM Congruent	cible-HFs PHYSIM Incongruent	cible-HFs PHYDIS Congruent	cible-HFs PHYDIS Incongruent	cible-BFs PHYSIM Congruent	cible-BFs PHYSIM Incongruent	cible-BFs PHYDIS Congruent	cible-BFs PHYDIS Incongruent
p(RC)	0.78 \pm 0.19	0.74 \pm 0.10	0.82 \pm 0.14	0.81 \pm 0.10	0.90 \pm 0.09	0.54 \pm 0.13	0.86 \pm 0.11	0.67 \pm 0.16
TRm (ms)	779 \pm 116	849 \pm 125	798 \pm 119	827 \pm 119	622 \pm 81	663 \pm 99	611 \pm 52	679 \pm 65
log(TR)m	6.62 \pm 0.15	6.70 \pm 0.15	6.64 \pm 0.15	6.67 \pm 0.14	6.40 \pm 0.12	6.46 \pm 0.14	6.39 \pm 0.08	6.49 \pm 0.09

Analyse de la proportion de réponses correctes

Les moyennes et écart-types de la proportion de réponses correctes pour chaque condition expérimentale sont reportés dans le Tableau 2.2. L'effet principal de la Fréquence spatiale cible n'est pas significatif ($\beta = 0.05$, $z = 0.31$, $p = .760$), et la Fréquence spatiale cible n'interagit pas avec le Temps de présentation ($\beta = -0.44$, $z = -1.69$, $p = .091$). Le modèle d'interaction montre que la Fréquence spatiale cible interagit avec la Congruence ($\beta = 1.01$, $z = 9.60$, $p < .001$). Les tests des effets simples révèlent que l'effet d'interférence est significatif à la fois dans la condition cible-HFs (congruent = 0.76 ± 0.17 , incongruent = 0.69 ± 0.14 ; $\beta = -0.38$, OR = 1.47, $z = -5.30$, $p < .001$) et dans la condition cible-BFs (congruent = 0.85 ± 0.10 , incongruent = 0.59 ± 0.12 ; $\beta = -1.39$, OR = 4.00, $z = -18.04$, $p < .001$). À l'inverse de nos prédictions, l'effet d'interférence est plus grand dans la condition cible-BFs (interférence des HFs, OR = 4.00) que dans la condition cible-HFs (interférence des BFs, OR = 1.47).

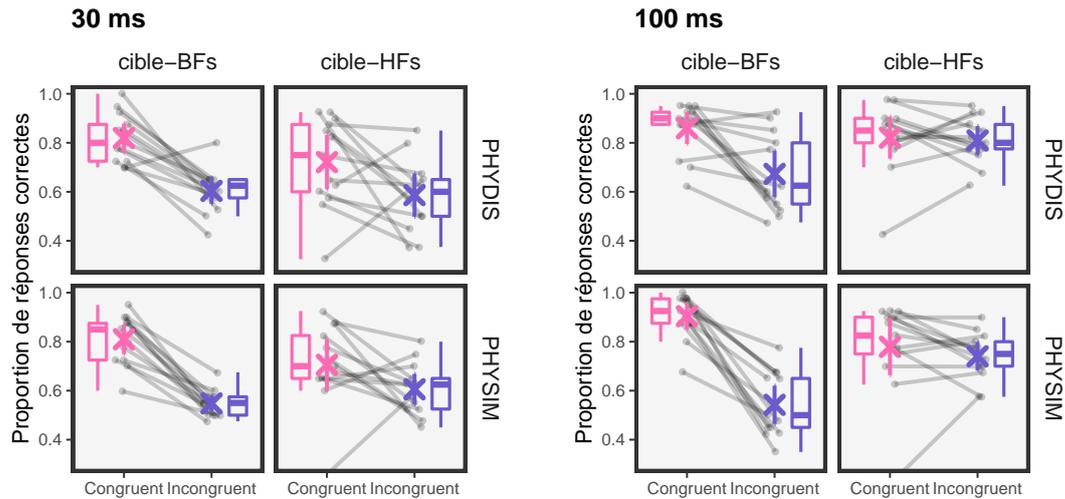


Figure 2.7: Résultats de l'Expérience 1 (Session 2 - Scènes hybrides) : Proportion de réponses correctes en fonction du Temps de présentation (30 ms ; 100 ms), de la Fréquence spatiale cible (cible-BFs ; cible-HFs), de la Congruence sémantique (congruent ; incongruent) et de la Similarité physique (similarité physique, PHYSIM ; dissimilarité physique, PHYDIS) entre les deux scènes filtrées composant l'hybride. Les \times représentent les moyennes et les barres d'erreur qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant.

Nous avons ensuite testé l'effet du Temps de présentation sur l'interaction entre la Congruence sémantique et la Fréquence spatiale cible. Le modèle de double interaction n'a pas pu converger. Nous avons donc analysé cet effet d'interaction avec un modèle linéaire classique (sans prise en compte des effets aléatoires). Cette analyse montre que l'interaction entre le Temps de présentation, la Fréquence spatiale cible et la Congruence sémantique n'est pas significative ($F(1,48) = 2.36, p = .131$).

Afin d'analyser l'effet de la similarité physique sur l'effet d'interférence, nous avons testé l'interaction entre la Similarité physique et la Congruence d'abord globalement, puis dans chaque condition croisée de Temps de présentation et de Fréquence spatiale cible. Le modèle d'interaction globale montre l'effet de Congruence est plus grand lorsque la scène distractive de l'image hybride est physiquement similaire à la scène cible, que lorsqu'elle est physiquement dissimilaire ($\beta = -0.23, z = -2.26, p = .024$). Dans la condition cible-HFs, l'effet d'interaction entre la Congruence et la Similarité physique n'est significatif ni lorsque le Temps de présentation est de 30 ms ($\beta = 0.16, z = 0.85, p = .396$), ni lorsqu'il est de 100 ms ($\beta = -0.14, z = -0.65, p < .516$). Dans la condition cible-BFs, l'effet d'interaction entre la Congruence sémantique et la Similarité physique n'est pas significatif lorsque le Temps de présentation est de 30 ms ($\beta = -0.17, z = -0.82, p = .411$), mais cet effet est significatif lorsque le Temps de présentation est de 100 ms ($\beta = -1.02, z = -4.31, p < .001$). Dans cette condition, l'effet d'interférence est plus fort

lorsque les scènes composant l'hybride sont physiquement similaires ($\beta = -2.13$, OR = 8.40, $z = -12.09$, $p < .001$) que lorsqu'elles sont physiquement dissimilaires ($\beta = -1.14$, OR = 3.13, $z = -7.10$, $p < .001$).

Analyse des temps de réponses correctes

Les moyennes et écart-types des TR moyens et des $\log(\text{TR})$ moyens pour chaque condition expérimentale sont reportés dans le Tableau 2.2. L'analyse montre un effet significatif de la Fréquence spatiale cible se traduisant par des TRm plus courts lorsque la scène cible dans l'hybride est en BFs que lorsqu'elle est en HFes (cible-BFs = 654 ms \pm 97, cible-HFs = 816 ms \pm 141; $\beta = 0.21$, %change = 23.84%, $t(49.87) = 5.18$, $p < .001$). Le modèle d'interaction montre que la Fréquence spatiale cible interagit avec la Congruence ($\beta = 0.05$, $t(5807) = 3.62$, $p < .001$). Le modèle d'effet simple révèle que l'effet d'interférence est significatif dans la condition cible-HFs (congruent = 802 \pm 147 ms, incongruent = 830 \pm 135 ms; $\beta = -0.04$, %change = 3.78%, $t(2914) = -3.99$, $p < .001$). Dans la condition cible-BFs, le modèle d'effet simple n'a pas pu converger, nous avons donc utilisé un modèle linéaire sans effet aléatoire. Cette analyse montre que dans cette condition, l'effet d'interférence est également significatif (congruent = 624 \pm 76 ms, incongruent = 683 \pm 108 ms; $\beta = 0.08$, %change = 8.58%, $t(25) = 6.50$, $p < .001$). Comme pour la proportion de réponses correctes, l'effet d'interférence est plus grand dans la condition cible-BFs (interférence des HFes, %change = 8.58%) que dans la condition cible-HFs (interférence des BFs, %change = 3.78%).

Nous avons ensuite testé l'effet du Temps de présentation sur l'interaction entre la Congruence sémantique et la Fréquence spatiale cible. Le modèle de double interaction ne montre pas d'effet significatif ($\beta = -0.00$, $t(5805) = -1.68$, $p = .093$).

Enfin, afin d'analyser l'effet de la similarité physique sur l'effet d'interférence, nous avons testé l'interaction entre la Similarité physique et la Congruence d'abord globalement puis dans chaque condition croisée de Temps de présentation et de Fréquence spatiale cible. L'effet d'interaction global entre la Similarité physique et la Congruence n'est pas significatif ($\beta = -0.01$, $t(5802) = -0.67$, $p = .506$). Dans la condition cible-HFs, l'interaction entre la Congruence et la Similarité physique n'est significative ni lorsque le Temps de présentation est de 30 ms ($\beta = 0.03$, $t(1323.13) = 1.14$, $p = .255$), ni lorsque que le Temps de présentation est de 100 ms ($\beta = -0.05$, $t(1583.15) = -1.81$, $p = .071$). De la même façon, dans la condition cible-BFs, l'interaction entre la Congruence et la Similarité physique n'est significative ni lorsque le Temps de présentation est de 30 ms ($\beta = -0.04$, $t(1391.14) = -1.55$, $p = .121$), ni lorsqu'il est de 100 ms ($\beta = 0.02$, $t(1496.10) = 0.99$, $p = .320$).

2.4. Discussion

Cette expérience avait pour objectif d'étudier comment les informations issues du traitement des basses et hautes fréquences spatiales interagissent lors de la catégorisation de scènes. Dans une première session expérimentale, les participants devaient catégoriser

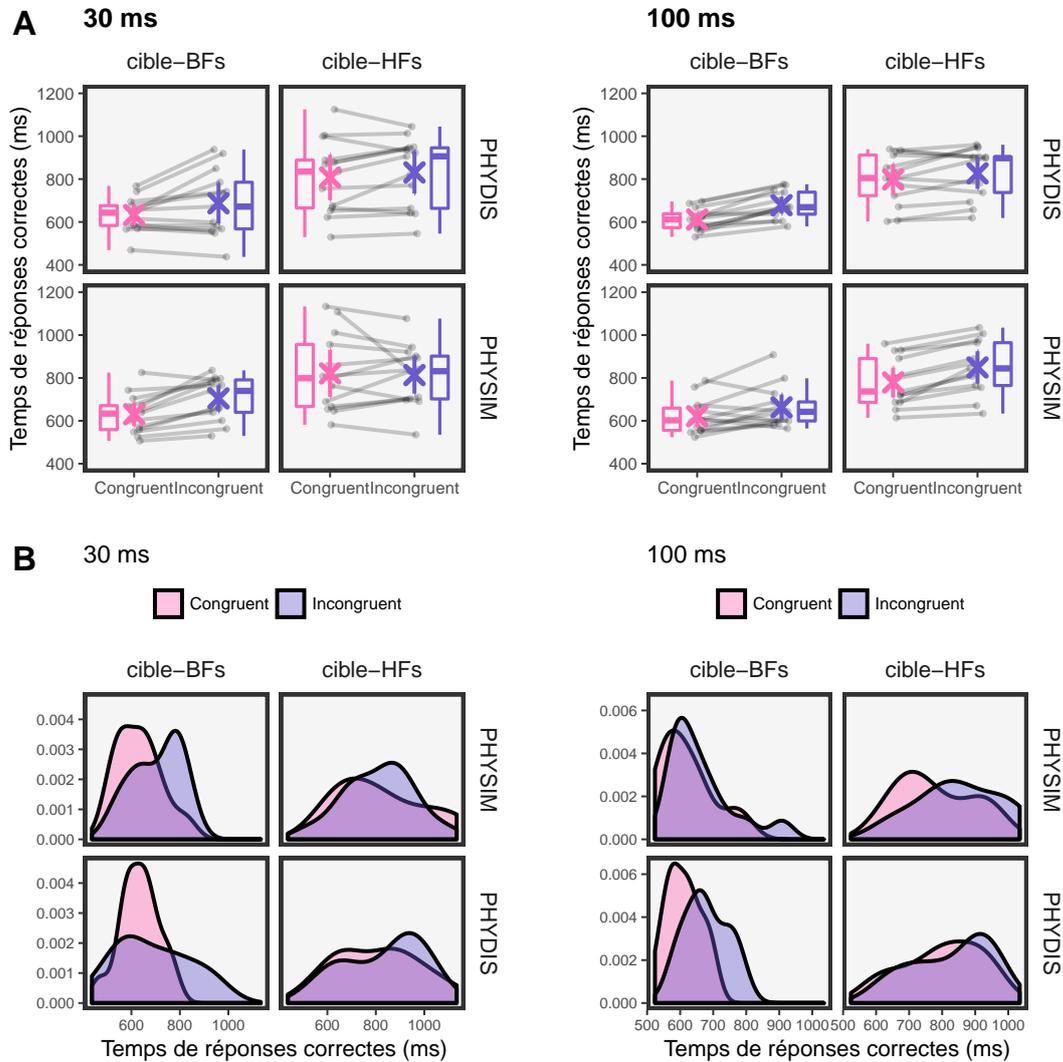


Figure 2.8: Résultats de l'Expérience 1 (Session 2 - Scènes hybrides)
A. Temps de réponses correctes en fonction de la Fréquence spatiale cible (cible-BFs; cible-HFs), du Temps de présentation (30 ms; 100 ms), de la Congruence sémantique (congruent; incongruente) et de la Similarité physique (similarité physique, PHYSIM; dissimilarité physique, PHYDIS) entre les deux scènes filtrées composant l'hybride. Les \times représentent les moyennes et les barres d'erreur qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **B.** Distributions comparées des temps de réponses correctes pour les hybrides congruents et pour les hybrides incongruents.

des scènes filtrées en BFs et HFs présentées séparément. Dans une seconde session expérimentale, nous avons présenté des scènes hybrides composées d'une scène BFs superposée à une scène HFs. Les participants devaient catégoriser leur contenu en BFs ou en HFs. Nous avons manipulé le temps de présentation des stimuli, qui étaient présentés pendant 30 ou 100 ms. Pour les stimuli hybrides, nous avons aussi manipulé la congruence sémantique et la similarité physique entre les deux scènes filtrées des hybrides. Nous avons considéré l'effet d'interférence (i.e. de moins bonnes performances lorsque les scènes de l'hybride sont sémantiquement congruentes que lorsqu'elles sont sémantiquement incongruentes) comme étant la signature de l'intégration des informations contenues dans les BFs et HFs. Plus précisément, ce paradigme expérimental nous a permis d'observer comment l'information issue d'une bande de fréquence spatiale qui n'est pas supposée être traitée explicitement par le participant influence le traitement d'une autre bande de fréquences spatiales. Il nous a également permis d'estimer le poids relatif des BFs et HFs lors de la catégorisation de scènes. En premier lieu, il est important de noter que les scènes filtrées BFs et HFs présentées séparément (Session 1) étaient facilement catégorisées par les participants (>80% de catégorisation correctes en moyenne). Cela nous a permis de nous assurer que les BFs et les HFs des images hybrides ont pu être traitées, même avec un temps de présentation très court (30 ms). Par ailleurs, les participants qui formaient les deux groupes de la Session 2 se comportaient de manière similaire lorsqu'ils catégorisaient les scènes filtrées BFs et HFs présentées séparément (voir Figure 2.6).

Avantage des BFs lors de la catégorisation rapide de scènes

En nous basant sur le modèle FP, nous avons d'abord fait l'hypothèse d'une précé-
dence temporelle du traitement des BFs par rapport au traitement des HFs. Les résultats de la Session 1 sont en accord avec cette hypothèse. Les analyses suggèrent que la catégorisation des scènes BFs est plus précise (OR = 1.60) et plus rapide (%change = 5.00%) que celle des HFs. Les modèles statistiques ajustés aux résultats de la Session 2 avec les scènes hybrides montrent que la catégorisation du composant cible en BFs dans une scène hybride est plus rapide que la catégorisation du composant cible en HFs (%change = 23.84%). La confrontation des résultats des deux sessions expérimentales montre donc que les performances de catégorisation sont toujours meilleures pour catégoriser les scènes sur la base des BFs que sur la base des HFs, mais traduit aussi un effet décuplé de la fréquence spatiale sur le temps de réponse lors de la catégorisation des scènes dans les images hybrides par rapport à la catégorisation de scènes filtrées individuelles (%change = 23.84% vs 5.00%, respectivement). En outre, on remarque que les temps de réponses correctes sont de l'ordre de 600 ms pour catégoriser les scènes filtrées en BFs et en HFs (Session 1) et la scène en BFs dans les hybrides (Session 2). Par contre, les temps de réponses correctes sont plus longs pour catégoriser la scène HFs dans les hybrides, de l'ordre de 800 ms. L'ensemble de ces résultats suggère que l'information contenue dans les HFs n'est pas la plus facile et rapide à catégoriser, et ce surtout lorsqu'elle n'est pas isolée, et donc moins saillante. La catégorisation rapide de scènes est au contraire plus efficace lorsqu'elle est basée sur les BFs. Ces résultats

appuient donc l'hypothèse d'une reconnaissance précoce basée sur l'analyse rapide des BFs de la scène, telle que le suppose le modèle FP.

Il est toutefois important de noter que le contraste de luminance n'était pas égalisé entre les scènes BFs et les scènes HFes dans notre expérience. Comme nous l'avons vu dans la partie 2.1.2 du Chapitre 1, le contraste de luminance des scènes naturelles décroît lorsque la fréquence spatiale augmente (Field 1987). Nos scènes BFs sont donc caractérisées par un plus grand contraste que les scènes HFes (RMS moyen \pm écart-type des scènes BFs 'naturelles' : 0.24 ± 0.06 ; scènes BFs 'artificielles' : 0.24 ± 0.05 ; scènes HFes 'naturelles' : 0.06 ± 0.02 ; scènes HFes 'artificielles' : 0.07 ± 0.02 ; ces valeurs représentent l'écart-type de la luminance pour une échelle allant de 0 à 1). On ne peut donc pas totalement exclure que l'avantage des BFs pour la catégorisation rapide observé ici soit en partie dû à une différence de contraste entre les scènes BFs et les scènes HFes. La contribution des différences de contraste entre les fréquences spatiales dans le contexte d'une stratégie d'analyse *coarse-to-fine* de scènes a fait l'objet d'une étude réalisée par Kauffmann et al. (2015a). Ces auteurs ont utilisé un paradigme expérimental permettant d'examiner séparément l'effet de la fréquence spatiale, celui du contraste et de leur combinaison. Leurs stimuli étaient des séquences d'images composées de six versions filtrées d'une même scène et assemblées en allant des BFs aux HFes (séquences *coarse-to-fine*), ou assemblées en allant des HFes aux BFs (séquences *fine-to-coarse*). Les résultats de cette étude montrent que les séquences *coarse-to-fine* sont mieux catégorisées que les séquences *fine-to-coarse*, et que même si l'avantage est en partie attribuable aux différences de contraste entre les fréquences spatiales, il repose principalement sur les différences en termes de fréquences spatiales. Nous pouvons donc être assez confiants quant à la contribution du contenu en fréquences spatiales, relativement à celui en contraste, dans les effets que nous observons ici.

Effet d'interférence : Influences réciproques des BFs et des HFes

La Session 2 de l'expérience (scènes hybrides) a permis de répliquer les résultats précédemment obtenus avec le même paradigme expérimental (Kauffmann et al. 2015c; Mu & Li 2013). En effet, nous observons un effet d'interférence des BFs distractives sur la catégorisation des HFes (condition cible-HFes). Les modèles statistiques estiment que les chances de catégoriser correctement la scène en HFes dans l'hybride sont plus élevées si la scène distractive en BFs est sémantiquement congruente qu'incongruente (OR = 1.47), et que cette catégorisation est plus rapide ($\%change = 3.78\%$). Comme nous l'avons prédit sur la base du modèle FP, les BFs seraient analysées très rapidement et permettraient d'activer des prédictions sur le contenu de la scène. Ces prédictions permettraient ensuite de guider le traitement ascendant des HFes dans le système de reconnaissance. Ainsi, dans une situation de d'incongruence sémantique entre les HFes et les BFs, le traitement automatique des BFs entraînerait parfois des prédictions erronées qui génèraient la catégorisation basée sur les HFes.

Notre expérience montre également que la catégorisation de la scène en BFs, bien

que plus rapide que celle de la scène en HFs, subit une forte interférence de la présence d'information en HFs (condition cible-BFs). Les modèles statistiques estiment que les chances de catégoriser correctement une scène en BFs sont plus élevées si la scène distractive en HFs est sémantiquement congruente qu'incongruente ($OR = 4.00$), et que cette catégorisation est plus rapide ($\%change = 8.58\%$). De manière surprenante et à l'inverse de ce que nous avons prédit, cet effet d'interférence des HFs sur la catégorisation des BFs est très fort et robuste (on peut remarquer sur les Figures 2.7 et 2.8 que l'interférence est présente chez tous les participants de la condition cible-BFs, ou presque). Il est plus grand que l'effet d'interférence des BFs sur la catégorisation des HFs, suggérant que l'information issue des HFs interfère plus fortement sur la catégorisation basée sur les BFs que l'inverse. Nous n'observons pas de modulation de ces effets d'interférence par le temps de présentation des scènes hybrides. Les résultats indiquent plutôt que l'effet d'interférence des HFs est plus fort que celui des BFs, quel que soit le temps de présentation. Ce résultat contredit partiellement nos hypothèses puisque nous nous attendions à ce que le poids de l'information sémantique contenue dans les BFs soit prépondérant avec un temps de présentation court (i.e. 30 ms), et qu'éventuellement, l'information sémantique contenue dans les HFs gagne de l'importance avec un temps de présentation plus long (i.e. 100 ms). Comment alors expliquer une si forte influence de l'information en HFs sur la catégorisation des scènes en BFs alors que les temps de présentation sont extrêmement courts ?

Notre interprétation est qu'en cas d'incohérence entre les informations sémantiques portées par les BFs et les HFs, comme c'est le cas avec les scènes hybrides, il pourrait être difficile et même contre-productif d'ignorer l'information accumulée en HFs au profit des seules représentations grossières issues de l'analyse des BFs. Cette tâche de catégorisation des stimuli hybrides pourrait donc instaurer une situation bien particulière que nous n'avions pas prévue. La stratégie principalement utilisée pour traiter des images hybrides serait alors celle d'une analyse minutieuse de sorte que (i) l'information détaillée portée par les HFs, même si longue à traiter, serait celle recherchée en priorité par le système de reconnaissance du modèle FP lors de la tâche et (ii) les traitements ascendants seraient favorisés, réduisant alors le poids des prédictions par rapport à des situations plus naturelles. Autrement dit, dans cette tâche comportementale difficile, où la possibilité d'incongruence sémantique entre des informations visuelles superposées est évidente, le système de reconnaissance pourrait réduire le poids des prédictions et accorder plus de poids à une analyse minutieuse des informations ascendantes, et notamment celles en HFs qui fournissent des indices de plus grande précision.

Lorsque les participants doivent catégoriser la scène en HFs, cette stratégie d'analyse minutieuse serait donc profitable, malgré le fait qu'un certain nombre de catégorisations soient tout de même influencées par l'analyse des BFs et témoignent d'erreurs de prédiction (i.e. effet relativement plus petit de l'interférence des BFs). Par contre, lorsque les participants doivent catégoriser la scène en BFs, les réponses des participants pourraient refléter une alternance entre les besoins de la tâche (porter l'attention sur les

BFs qui permettraient une catégorisation très rapide de la scène) et la stratégie involontaire déclenchée par la nature ambivalente du stimulus (i.e. analyse privilégiée des HFs). L'importante intégration ascendante des HFs viendrait donc interférer sur le traitement attentionnel dirigé sur les BFs.

Effet de la similarité physique

Indépendamment de la congruence sémantique, nous avons également manipulé la similarité physique entre les scènes en BFs et en HFs composant les scènes hybrides. Pour rappel, [Mu et Li \(2013\)](#) avaient observé une augmentation de l'effet d'interférence (sur la proportion de réponses correctes, mais pas sur les temps de réponse) lorsque les deux scènes étaient physiquement similaires, par rapport à lorsqu'elles étaient physiquement dissimilaires. Cet effet d'interaction entre la congruence sémantique et la similarité physique n'a pas été répliqué par [Kauffmann et al. \(2015c\)](#). Dans notre expérience, nous observons une interaction globale entre la similarité physique et la congruence sémantique, indiquant que le recouvrement spatial et spectral entre les BFs et les HFs pourrait jouer un rôle dans leur intégration lors de la catégorisation de scènes. Cependant, il est difficile, sur la base de nos résultats, de savoir plus exactement comment ces deux types d'information (physique/bas niveau et sémantique/haut niveau) interagissent. En effet, nous ne répliquons pas l'effet observé par [Mu et Li \(2013\)](#). Les BFs n'interfèrent pas plus fortement sur les HFs lorsqu'elles sont physiquement similaires que lorsqu'elles sont physiquement dissimilaires (condition cible-HFs). Par contre, les HFs interfèrent plus fortement sur les BFs lorsqu'elles sont physiquement similaires ($OR = 8.40$) que lorsqu'elles sont physiquement dissimilaires ($OR = 3.13$) - cet effet d'interaction étant observé seulement lorsque l'image hybride est présentée 100 ms, mais pas lorsqu'elle est présentée pendant 30 ms. Ce résultat est difficile à interpréter en tant qu'utilisation de l'information physique dans les mécanismes prédictifs, puisque comme nous l'avons vu, l'interférence des HFs pourrait être liée à un traitement ascendant persistant des HFs lors de la tâche. Ainsi, cet effet d'interaction dans la condition cible-BFs pourrait plutôt indiquer que les HFs gagnent en saillance avec le temps de présentation, comme le prévoirait le modèle FP.

Conclusion

En résumé, l'interférence des BFs sur la catégorisation des HFs refléterait la conséquence d'une erreur de prédiction, alors que la forte interférence des HFs sur les BFs pourrait refléter plutôt l'intégration ascendante persistante des HFs qui vient à l'encontre des représentations rapides issues de l'analyse des BFs. Concernant la nature de l'influence prédictive, nos résultats ne montrent pas d'influence de la similarité physique dans l'interférence des BFs sur les HFs, ce qui suggère que les prédictions issues du traitement des BFs sont de nature plutôt sémantique. Ce résultat est tout de même à confronter avec ceux de [Mu et Li \(2013\)](#) qui suggèrent au contraire que les prédictions véhiculent aussi une information de bas niveau (i.e. liée à la configuration spatiale et au spectre d'amplitude). De futures études sont nécessaires pour clarifier la nature du signal

prédictif. Quoi qu'il en soit, les résultats de l'Expérience 1 montrent que les HF, aussi bien que les BF, peuvent être utilisées pour guider une tâche de catégorisation, et leur intégration pourrait dépendre d'interactions complexes entre la nature de l'information perceptive et les caractéristiques de la tâche, qui impliquent ici la présence d'un conflit sémantique apparent au sein du stimulus. Par rapport aux études précédentes (Kauffmann et al. 2015c; Mu & Li 2013), notre expérience nous incite à insister sur le fait que les scènes hybrides sont des stimuli très peu naturels. Au vu de l'interprétation que nous avons des résultats obtenus, nous concluons que les stimuli hybrides n'ont qu'un intérêt limité pour étudier l'intégration des fréquences spatiales, notamment parce que les résultats sont moins facilement généralisables à des conditions naturelles où la superposition de fréquences spatiales non congruentes n'existe a priori pas. Dans les Expériences 2 et 3, nous avons pour objectif de tester le modèle FP-peri, c'est-à-dire l'hypothèse que les prédictions visuelles peuvent être activées sur la base de l'information disponible en vision périphérique. Nous avons de nouveau utilisé un paradigme d'interférence et nous avons pu nous affranchir du problème de superposition des fréquences spatiales des images hybrides en présentant une scène en vision périphérique et une scène en vision centrale. Nous avons ainsi pu tester l'interférence d'une scène périphérique sur la catégorisation d'une scène centrale.

3. Expériences 2 et 3. Influence d'une scène en vision périphérique sur la catégorisation de scènes en vision centrale

3.1. Objectifs et hypothèses

Dans cette partie expérimentale, nous avons pour objectif d'étudier si les informations disponibles en vision périphérique sont susceptibles de moduler le traitement de l'information visuelle présente en vision centrale. Nous avons couplé mesures comportementales (Expérience 2 et Expérience 3) et mesures d'IRM fonctionnelle (Expérience 3). Nous avons adapté les stimuli et le protocole expérimental de l'Expérience 1, à savoir que nous avons présenté les deux scènes qui composaient une image hybride dans l'Expérience 1 simultanément sur l'axe horizontal, l'une en vision centrale et l'autre en vision périphérique, dans le champ visuel droit ou gauche. Contrairement à l'Expérience 1, les deux scènes n'étaient pas filtrées, mais du fait de la résolution spatiale non homogène de l'information visuelle à travers le champ visuel, nous avons considéré que la présentation des scènes en vision périphérique agirait comme un filtre passe-bas naturel. Les deux scènes pouvaient appartenir à la même catégorie (congruence sémantique) ou à des catégories différentes (incongruence sémantique). Les participants devaient catégoriser la scène présentée en vision centrale selon qu'il s'agissait d'une scène appartenant à la catégorie artificielle ou à la catégorie naturelle, tout en ignorant la scène présentée en vision périphérique.

Dans le cadre du modèle FP-péri, notre hypothèse était que lors de la catégorisation de la scène en vision centrale, le traitement rapide des BFs de la scène au centre et celui des BFs de la scène en périphérie permettrait l'émergence d'une première représentation rudimentaire de type *gist*. Cette représentation pourrait ainsi rapidement activer les mécanismes de prédictions, lesquels viendraient ensuite contraindre le traitement ultérieur des HFes extraites en vision centrale. Il est important de noter que le présent paradigme ne nous permet pas de tester l'influence de prédictions issues de la vision centrale, puisqu'ici les contenus en BFes et en HFes de la scène centrale sont nécessairement de la même catégorie. Par contre, si des prédictions émanant de la vision périphérique viennent guider le traitement de l'information présente en vision centrale, nous nous attendions à observer un effet d'interférence de la scène périphérique sur la catégorisation de la scène centrale. Dans ce contexte, les deux expériences que nous avons réalisées avaient pour but de répondre à plusieurs questions. En guise de guide de lecture, ces questions et les hypothèses qui y sont associées sont détaillées ci-après. Le dernier point concerne les bases cérébrales, et donc seulement l'Expérience 3 (IRMf).

3. Expériences 2 et 3. Influence d'une scène en vision périphérique sur la catégorisation de scènes en vision centrale

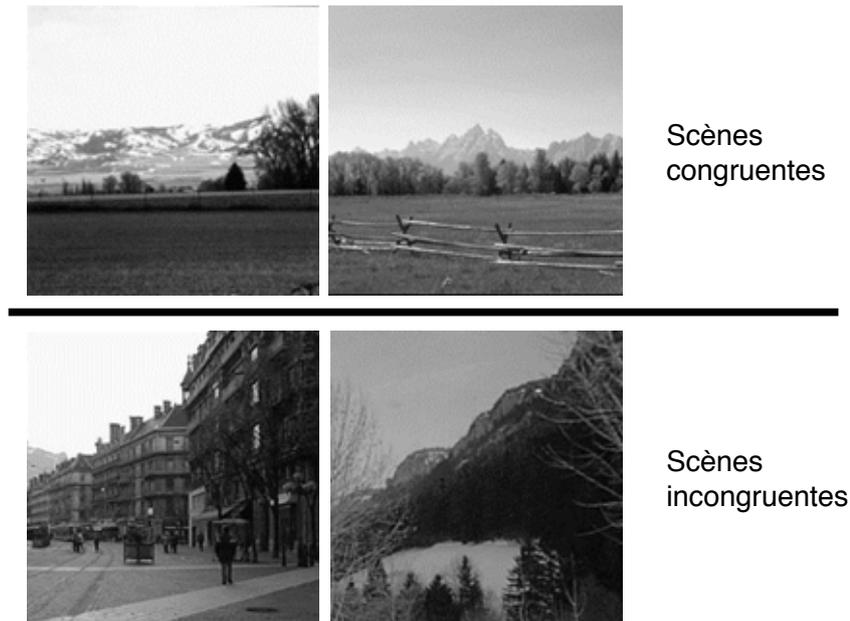


Figure 2.9: Exemples de stimuli utilisés dans l'Expérience 2. Une scène était présentée en vision centrale tandis que l'autre était présentée en vision périphérique sur l'axe horizontal. Les deux scènes étaient congruentes (deux scènes artificielles ou deux scènes naturelles) ou incongruentes (une scène artificielle et une scène naturelle).

3.1.1. Test d'une influence de l'information en BFs extraite en vision périphérique sur la catégorisation de scènes en vision centrale

L'objectif principal de ces expériences était donc de déterminer si l'information en BFs disponible en vision périphérique peut être intégrée à l'information traitée en vision centrale. L'hypothèse inverse serait que seules les informations en BFs et HFes qui se recouvrent spatialement en vision centrale sont intégrées. Dans le cadre du modèle FP-peri, nous avons fait l'hypothèse que la confrontation des informations contenues dans les BFs extraites en vision centrale et de celles contenues dans les BFs incongruentes extraites en vision périphérique mènent, au sein du système de gestion des prédictions, à l'élaboration d'un signal prédictif de mauvaise qualité. En arrivant sur le système de reconnaissance, ce signal pourrait ainsi perturber la catégorisation. Cela se traduirait sur la mesure comportementale par un nombre d'erreurs de catégorisation de la scène centrale plus important et des temps de réponse plus longs lorsque la scène en vision périphérique n'appartient pas à la même catégorie sémantique que la scène cible à catégoriser en vision centrale (i.e. un effet d'interférence).

3.1.2. Nature de l'influence prédictive périphérique : Influence de la similarité physique

Comme dans l'Expérience 1 avec les scènes hybrides, nous avons manipulé la similarité physique entre les deux scènes centrale et périphérique. Ainsi, les scènes centrale et

périphérique sémantiquement congruentes pouvaient être soit physiquement similaires (configuration spatiale et spectre d’amplitude similaires), soit physiquement dissimilaires. De même, les scènes centrale et périphérique sémantiquement incongruentes pouvaient être soit physiquement similaires (configuration spatiale et spectre d’amplitude similaires), soit physiquement dissimilaires. Cette manipulation permettait d’évaluer si les prédictions issues de l’analyse de la scène périphérique sont purement liées à un contenu sémantique, ou si au contraire des propriétés visuelles de plus bas niveau sont préservées dans le signal prédictif pour contraindre l’analyse de la scène centrale. Même si les résultats de l’Expérience 1 n’ont pas permis d’étayer cette hypothèse, nous nous attendions à ce que l’interférence du signal prédictif préserve des propriétés visuelles de bas niveau, comme par exemple une information sur le contenu en fréquences spatiales et en orientations. Nous nous attendions donc à un effet d’interaction entre la congruence sémantique et la similarité physique des scènes, à savoir une plus grande interférence de la scène en vision périphérique dans le cas où les deux scènes sont physiquement similaires.

3.1.3. Bases cérébrales de l’interférence périphérique

Dans l’Expérience 3, nous nous sommes intéressés au réseau cérébral impliqué dans l’interférence de la scène périphérique sur la catégorisation de la scène centrale. Pour étudier les régions cérébrales impliquées dans l’intégration de l’information périphérique aux processus de reconnaissance visuelle en vision centrale, nous avons adapté le paradigme expérimental à l’IRMf et analysé le signal BOLD avec la méthode de soustraction cognitive. La tâche et les stimuli étaient les mêmes que pour l’Expérience 2. Au niveau cérébral, les modèles neurocognitifs de reconnaissance visuelle (Bar 2003; Bar et al. 2006; Kauffmann et al. 2015b, 2014; Peyrin et al. 2010) suggèrent que le premier signal disponible en BFs serait rapidement traité par le cortex frontal inférieur, et notamment le cortex orbitofrontal, et activerait des interprétations possibles de la catégorie du stimulus visuel. Ces prédictions influenceraient ensuite le traitement plus lent du signal en HF le long du cortex ventral occipito-temporal. Plus précisément, le signal prédit sur la base des BFs serait comparé au signal HF entrant, et l’erreur de prédiction serait retransmise vers les aires de plus haut niveau, permettant de mettre à jour les prédictions, et ainsi de suite jusqu’à atteindre un état stable. Ces va-et-vients entre les prédictions et les erreurs de prédictions résulteraient en une augmentation de l’activité cérébrale d’autant plus importante que les erreurs de prédictions seraient importantes (Fuster & Bressler 2015; Kauffmann et al. 2015c; Nobre et al. 1999). Nous nous attendions alors à ce que des erreurs de prédiction induites par le contenu sémantique, mais aussi par le contenu physique de la scène périphérique s’accompagne d’une plus forte activation des régions cérébrales impliquées dans le codage prédictif de l’information visuelle. Dans le cadre des modèles neurocognitifs de reconnaissance visuelle et compte tenu des résultats de Kauffmann et al. (2015c) obtenus avec des scènes hybrides présentées en vision centrale, nous avons donc prédit une activation plus importante du cortex frontal inférieur et plus exactement du cortex orbitofrontal lorsque les scènes centrale et périphérique

sont sémantiquement incongruentes que lorsqu'elles sont congruentes. Par ailleurs dans ces modèles, le cortex frontal inférieur influencerait le traitement de l'information en HFs dans le cortex ventral occipito-temporal (notamment le gyrus fusiforme et le cortex parahippocampique). Grâce à une analyse de connectivité fonctionnelle, [Kauffmann et al. \(2015c\)](#) ont en effet observé une plus grande connectivité entre le cortex frontal inférieur et le cortex occipito-temporal lorsque les scènes composant l'hybride étaient sémantiquement incongruentes (i.e. en situation d'interférence) que lorsqu'elles étaient sémantiquement congruentes, suggérant une plus grande communication entre ces régions en cas d'erreur de prédiction. Nous avons donc également prédit une activation plus importante des régions du cortex occipito-temporal lorsque les scènes centrale et périphérique sont sémantiquement incongruentes que lorsqu'elles sont congruentes.

Les deux expériences présentées ci-dessous sont complémentaires. Leurs résultats sont discutés conjointement dans la sous-partie [3.6](#).

3.2. Expérience 2 : Méthode

3.2.1. Participants

Vingt-six étudiants de psychologie de l'Université Grenoble Alpes (24 femmes, âge moyen \pm écart-type : 20 \pm 3 ans) droitiers ont participé à cette expérience. Tous rapportaient une acuité visuelle normale ou corrigée à la normale. Chaque participant a donné son consentement libre et éclairé en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine. À la fin de l'expérience, chaque participant a reçu un bon d'expérience de 0.5 point à valoir lors des examens universitaires.

3.2.2. Stimuli et procédure

Nous avons utilisé les mêmes scènes que celles qui composaient les hybrides de l'Expérience 1, soit 80 scènes appartenant à la catégorie 'naturelle' (e.g., montagnes, champs, plages) et 80 scènes appartenant à la catégorie 'artificielle' (e.g., rues, autoroutes). La manipulation des images a été réalisée dans l'environnement MATLAB. Ici, nous avons égalisé la luminance moyenne des scènes ainsi que leur contraste RMS à partir de la valeur moyenne de luminance (117 sur une échelle de 256 niveaux de gris) et de contraste (64 sur une échelle de 256 niveaux de gris) de l'ensemble des 160 scènes. Sur la base de l'Expérience 1, nous avons utilisé les 40 paires de scènes sémantiquement congruentes (deux scènes artificielles ou deux scènes naturelles) et 40 paires de scènes sémantiquement incongruentes (une scène artificielle et une scène naturelle). Pour chacune des conditions de congruence sémantique (congruent et incongruent), la moitié des paires était formée de scènes physiquement similaires (PHYSIM), et l'autre moitié était formée de scènes physiquement dissimilaires (PHYDIS). La similarité physique entre les scènes était évaluée de la même manière que pour celles composant les scènes hybrides dans l'Expérience 1. Pour résumer, nos stimuli étaient 20 paires de scènes congruent-PHYSIM, 20 paires

congruent-PHYDIS, 20 paires incongruent-PHYSIM, et 20 paires incongruent-PHYDIS (cf. Figure 2.10).

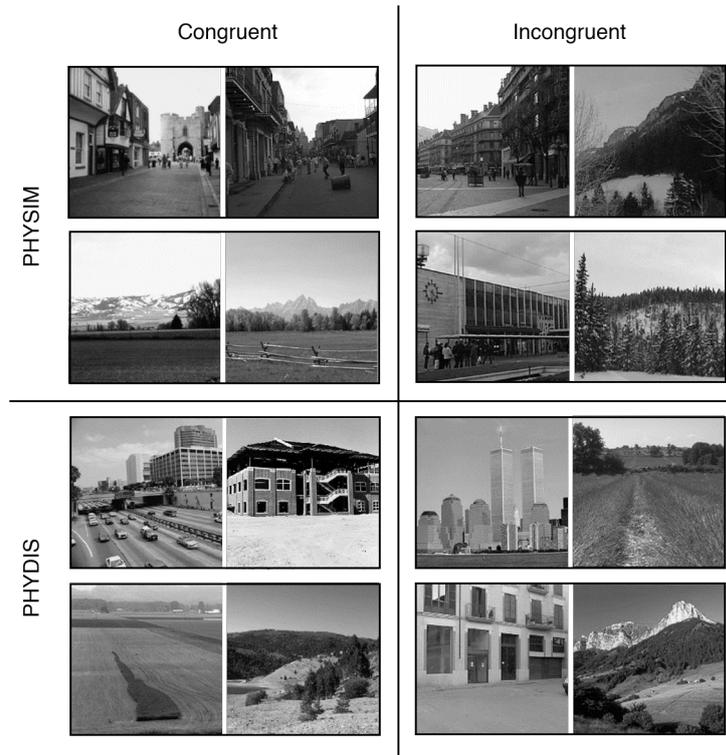


Figure 2.10: Types de stimuli utilisés dans l'Expérience 2. Une scène était présentée en vision centrale tandis que l'autre était présentée en vision périphérique sur l'axe horizontal, centrée à 7° d'excentricité. Les deux scènes de chaque paire pouvaient être (1) Congruentes et physiquement similaires (PHYSIM), (2) Congruentes et physiquement dissimilaires (PHYDIS), (3) Incongruentes et physiquement similaires (PHYSIM), ou (4) Incongruentes et physiquement dissimilaires (PHYDIS).

L'expérience était programmée avec E-Prime 2. Les stimuli (256×256 pixels) apparaissaient sur un écran d'ordinateur de 30 pouces (Dell UltraSharp) de résolution 2560×1600 pixels, avec un taux de rafraîchissement de 60 Hz. Les stimuli étaient affichés sur un fond gris uniforme de 117 sur une échelle de 256 niveaux de gris. Les participants étaient assis dans l'obscurité à une distance de 55 cm de l'écran. A cette distance, les stimuli couvraient $6 \times 6^\circ$ d'angle visuel. Pour maintenir une position centrée et la distance à l'écran, la tête des participants reposait sur une mentonnière positionnée dans l'axe sagittal médian. Chaque essai se déroulait de la manière suivante. Un point de fixation central apparaissait pendant 500 ms. Immédiatement après, une paire de scènes apparaissait pendant 100 ms, une scène cible étant présentée au centre de l'écran et l'autre de façon excentrée sur l'axe horizontal. La scène périphérique pouvait apparaître aléatoirement à droite ou à gauche de l'écran. Elle était centrée à 7° d'angle visuel, son bord interne se situait donc à 4° . Les participants disposaient ensuite de 2000 ms pour répondre (Figure 2.11). Ils avaient pour tâche de catégoriser le plus

précisément et le plus rapidement possible la scène centrale (scène cible) comme étant une scène appartenant soit à la catégorie 'naturelle', soit à la catégorie 'artificielle', tout en ignorant la scène présentée sur le côté (scène distractive). La variable de Similarité physique était manipulée en intra-sujet et en blocs, avec un bloc de scènes PHYSIM et un bloc de scènes PHYDIS. L'ordre des blocs était contrebalancé entre les participants. La variable de Congruence sémantique était manipulée en intra-sujet et randomisée au sein de chaque bloc de Similarité physique. Les participants répondaient aussi rapidement et précisément que possible à l'aide des touches 1 et 2 d'un clavier d'ordinateur, qui étaient contrebalancées entre les participants. À chaque essai, l'exactitude de la réponse et le temps de réponse étaient enregistrés. Chaque participant passait au total 320 essais (80 par condition expérimentale : congruent-PHYSIM, congruent-PHYDIS, incongruent-PHYSIM, incongruent-PHYDIS). L'expérience durait environ 20 minutes. La phase expérimentale était précédée d'une courte phase d'apprentissage afin que les participants s'habituent à la tâche et aux conditions de l'expérience (8 essais, avec des paires de scènes qui n'étaient pas présentées pendant les blocs expérimentaux).

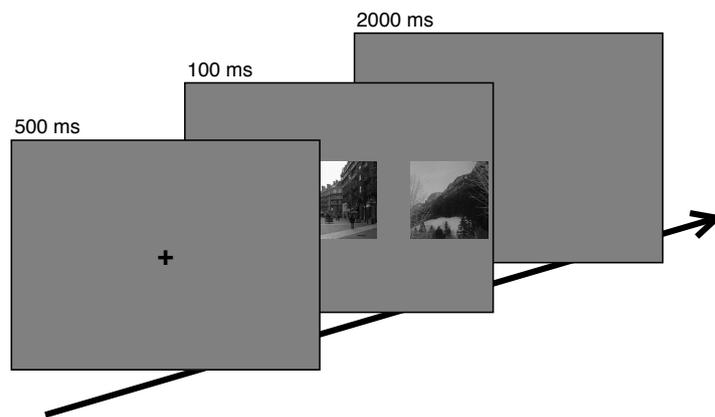


Figure 2.11: Procédure expérimentale de l'Expérience 2.

3.3. Expérience 2 : Résultats

Les effets fixes des modèles mixtes sont estimés pour la Congruence (congruent, incongruent) et la Similarité physique (PHYSIM, PHYDIS) entre les scènes, et leur interaction, et les effets aléatoires sont pris en compte pour l'ordonnée à l'origine des participants.

3.3.1. Analyse de la proportion de réponses correctes

Les moyennes et écart-types de la proportion de réponses correctes pour chaque condition expérimentale sont reportés dans le Tableau 2.3. L'analyse montre un effet significatif de la Congruence sémantique, la catégorisation de la scène centrale étant légèrement meilleure lorsque la scène périphérique est congruente que lorsqu'elle est incongruente (congruent = 0.96 ± 0.05 , incongruent = 0.95 ± 0.05 ; $\beta = -0.26$, OR =

1.29, $z = -2.32$, $p = .020$; Figure 2.12). La Similarité physique n'interagit pas avec la congruence sémantique ($\beta = 0.20$, $z = 0.89$, $p = .374$).

Tableau 2.3: Résultats de l'Expérience 2. Moyennes \pm écart-types de la proportion de réponses correctes (pRC), des temps de réponses correctes moyens en millisecondes (TRm), et des TR transformés (transformation logarithmique) moyens (log(TR)m) en fonction de la Congruence sémantique (congruent ; incongruent) et de la Similarité physique (similarité physique : PHYSIM ; dissimilarité physique : PHYDIS).

	PHYSIM Congruent	PHYSIM Incongruent	PHYDIS Congruent	PHYDIS Incongruent
pRC	0.96 \pm 0.05	0.95 \pm 0.05	0.96 \pm 0.05	0.96 \pm 0.05
TRm (ms)	643 \pm 104	668 \pm 107	610 \pm 99	613 \pm 86
log(TR)m	6.43 \pm 0.16	6.47 \pm 0.15	6.38 \pm 0.15	6.39 \pm 0.13

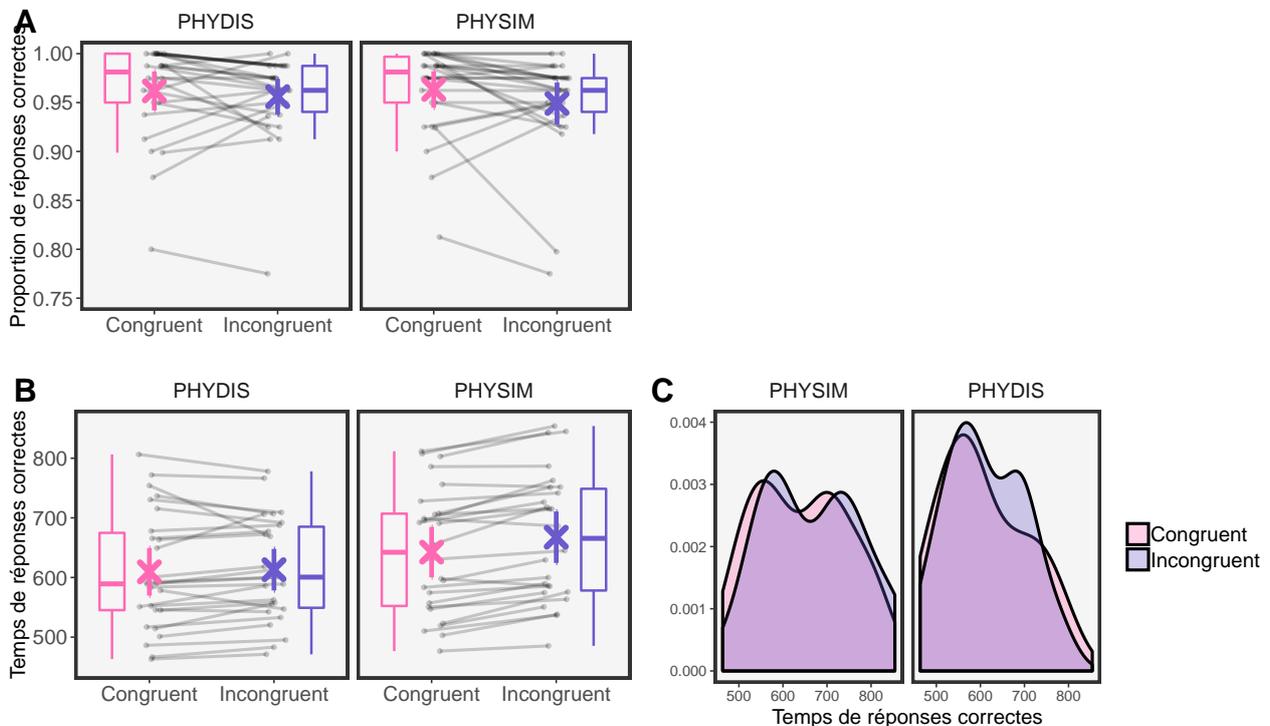


Figure 2.12: Résultats de l'Expérience 2. **A.** Proportion de réponses correctes et **B.** Temps de réponses correctes en fonction de la Congruence sémantique (congruent ; incongruent) et de la Similarité physique (similarité physique, PHYSIM ; dissimilarité physique, PHYDIS) entre les deux scènes. Les \times représentent les moyennes et les barres d'erreur qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **C.** Distributions comparées des temps de réponses correctes pour les conditions congruente et incongruente.

3.3.2. Analyse des temps de réponses correctes

Les moyennes et écart-types des TR_m et des log(TR)_m pour chaque condition expérimentale sont reportés dans le Tableau 2.3. L'analyse des temps de réponses correctes montre un effet principal significatif de la Congruence sémantique entre les deux scènes, la catégorisation de la scène centrale étant plus rapide lorsque la scène périphérique est congruente que lorsqu'elle est incongruente (congruent = 626 ± 102 , incongruent = 640 ± 100 ; $\beta = 0.02$, %change = 2.29%, $t(7693) = 4.72$, $p < .001$). Le modèle d'interaction montre que la Similarité physique interagit avec la Congruence sémantique ($\beta = 0.03$, $t(7691) = 2.97$, $p = .003$; Figure 2.12). Plus précisément, l'analyse des effets simples montre qu'il y a un effet d'interférence dans la condition de similarité physique ($\beta = 0.04$, %change = 3.78%, $t(3826) = 5.52$, $p < .001$), mais pas dans la condition de dissimilarité physique ($\beta = 0.009$, $t(3840) = 1.38$, $p = .168$).

3.4. Expérience 3 : Méthode

3.4.1. Participants

Quinze participants droitiers (dont 8 femmes; 23 ± 3 ans) ont participé à cette expérience. Tous rapportaient une acuité visuelle normale ou corrigée à la normale. Chaque participant a donné son consentement libre et éclairé avant de participer à l'expérience, qui a été menée en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine, et approuvée par un comité d'éthique (Comité de protection des personnes, Sud-Est V, ID/RCB : 2016-A00637-44).

3.4.2. Stimuli et procédure

Les stimuli utilisés étaient les mêmes que ceux de l'Expérience 2. L'expérience a été programmée avec E-Prime 2. Dans l'IRM, les stimuli étaient transmis sur un écran translucide situé derrière le scanner grâce à un vidéoprojecteur. Les participants pouvaient voir l'écran depuis l'intérieur du scanner par l'intermédiaire d'un miroir fixé sur l'antenne radio fréquence. Après avoir été installés dans le scanner, les participants effectuaient la tâche de catégorisation. Les participants avaient pour tâche de catégoriser (scène 'naturelle' ou scène 'artificielle') le plus précisément et le plus rapidement possible la scène centrale (scène cible), en ignorant la scène présentée sur le côté (scène distractive). Chaque essai se déroulait de la manière suivante. Un point de fixation central apparaissait pendant 500 ms sur un fond gris uniforme. Immédiatement après, une paire de scènes apparaissait pendant 100 ms, une scène étant présentée au centre de l'écran, en vision centrale, et l'autre latéralement et aléatoirement à droite ou à gauche dans le champ visuel périphérique. La taille angulaire des scènes était de $6 \times 6^\circ$. Les scènes périphériques étaient centrées à 6.75° d'angle visuel, leur bord interne se situait donc 3.75° . Les participants disposaient ensuite de 1900 ms pour répondre. Ils donnaient leur réponse de catégorisation à l'aide d'un boîtier de réponse placé à l'intérieur du scanner. Ils

avaient pour instruction de répondre aussi rapidement et précisément que possible avec deux boutons du boîtier de réponse, qui étaient contrebalancés entre les participants. À chaque essai, l'exactitude de la réponse et le temps de réponse étaient enregistrés. Avant l'expérience et en dehors du scanner, les participants réalisaient une courte phase d'apprentissage afin de s'habituer à la tâche. L'expérience comprenait deux scans fonctionnels conçus avec un paradigme IRMf événementiel de type pseudo-aléatoire. Bien qu'un paradigme de type bloc eut été plus robuste, la randomisation des essais des conditions congruente et incongruente est indispensable pour limiter au maximum les biais de demande. Un scan fonctionnel était composé de 160 événements (40 par condition expérimentale d'intérêt : congruent-PHYSIM, congruent-PHYDIS, incongruent-PHYSIM, incongruent-PHYDIS) et de 26 événements de repos (dont six à la fin du scan et pendant lesquels un point de fixation central était présenté sur un fond gris uniforme). Les paires étaient présentées selon deux ordres pseudo-aléatoires différents, répartis sur les deux scans fonctionnels. Pour chaque scan fonctionnel, 186 volumes fonctionnels étaient acquis (7 min 45 sec).

3.4.3. Acquisition des données IRMf

Les acquisitions ont été réalisées sur un scanner Philips Achieva 3T (Philips Medical Systems, Eugene, OR). Avant le premier essai de chaque session d'acquisition, cinq mesures du volume (*dummies scans*) étaient réalisées pour permettre la stabilisation du champ magnétique. Les acquisitions fonctionnelles (images pondérées en T2*) étaient ensuite réalisées en mode *echo planar imaging* (EPI) afin d'exploiter le contraste BOLD. Quarante-deux coupes adjacentes de 3 mm d'épaisseur, parallèles au plan bi-commissural, étaient acquises de manière séquentielle de bas en haut en incluant le cervelet (TR = 2500 ms, TE = 30 ms, angle de basculement = 82°, champ de vue = 240 × 240 × 126 mm, matrice d'acquisition = 80 × 80, matrice de reconstruction = 80 × 80, taille des voxels = 3 × 3 × 3 mm). Après les scans fonctionnels, un scan anatomique (images pondérées en T1, champ de vue = 256 × 224 × 175 mm, résolution = 1.33 × 1.750 × 1.37 mm, matrice d'acquisition = 192 × 115 × 128, matrice de reconstruction = 288 × 288 × 128) était acquis afin de pouvoir superposer les images fonctionnelles à une image anatomique de haute résolution spatiale et de pouvoir identifier les régions cérébrales par rapport à des atlas anatomiques de référence.

3.4.4. Traitement et analyse statistique des données d'IRMf

Les données d'IRMf ont été traitées en utilisant le programme *Statistical Parametric Mapping 12* (SPM, Wellcome Trust Centre for Neuroimaging, Londres, R.U.) dans MATLAB. Le traitement spatial des volumes fonctionnels de chaque participant a été réalisé selon les étapes suivantes. Nous avons tout d'abord effectué un *slice-timing* afin de corriger le décalage temporel d'acquisition entre les coupes composant un même volume fonctionnel. Nous avons ensuite corrigé les artefacts dus aux mouvements de la tête du participant (rotations et translations) en réalignant chaque volume fonctionnel

au premier volume du scan. À l'issue du réalignement, une image moyenne des images fonctionnelles a été créée. Nous avons alors appliqué une coregistration qui permet de faire correspondre les orientations géométriques de l'image anatomique à la moyenne des images fonctionnelles. L'image anatomique a été ensuite normalisée dans l'espace MNI (Montreal Neurological Institute) à l'aide d'un template anatomique, puis les images fonctionnelles ont été normalisées à l'aide des paramètres calculés lors de la normalisation de l'image anatomique. Enfin, les images fonctionnelles ont été lissées par un filtre gaussien afin d'augmenter le rapport signal sur bruit.

Les données fonctionnelles ont ensuite été modélisées avec un modèle linéaire général (Friston et al. 1995). Pour chaque participant, nous avons créé une matrice de dessin incluant quatre régresseurs correspondant aux différentes conditions expérimentales (congruent-PHYSIM, incongruent-PHYSIM, congruent-PHYDIS, incongruent-PHYDIS). Pour chaque régresseur, le moment de l'apparition (l'*onset*) de chaque paire de scènes a été convolué avec la fonction de réponse hémodynamique canonique. Les paramètres de mouvement (trois rotations et trois translations) ont également été modélisés dans la matrice de dessin comme régresseurs de non intérêt. Nous avons ensuite généré des cartes d'activation pour chaque participant à partir des contrastes entre les valeurs des paramètres estimés pour les différentes conditions expérimentales, et nous les avons ensuite entrées dans une analyse de groupe à effets aléatoires, à l'aide de tests de Student pour échantillons appariés. Nous avons retenu les clusters d'activation sur la base de l'intensité de la réponse individuelle des voxels (à un seuil de 0.05 avec une correction du *Family-Wise Error rate* - FWE - au niveau du cluster) et de l'étendue spatiale de l'activation (clusters d'au moins 10 voxels).

3.5. Expérience 3 : Résultats

Les effets fixes des modèles mixtes sont estimés pour la Congruence sémantique (congruent, incongruent) et la Similarité physique (PHYSIM, PHYDIS) entre les scènes, et les effets aléatoires sont pris en compte pour l'ordonnée à l'origine des participants.

Tableau 2.4: Résultats de l'Expérience 3. Moyennes \pm écart-types de la proportion de réponses correctes (pRC), des temps de réponses correctes moyens en millisecondes (TRm), et des TR transformés (transformation logarithmique) moyens (log(TR)m) en fonction de la Congruence sémantique (congruent ; incongruent) et de la Similarité physique (similarité physique : PHYSIM : dissimilarité physique : PHYDIS).

	PHYSIM Congruent	PHYSIM Incongruent	PHYDIS Congruent	PHYDIS Incongruent
pRC	0.95 \pm 0.05	0.93 \pm 0.06	0.96 \pm 0.04	0.97 \pm 0.03
TRm (ms)	616 \pm 80	638 \pm 93	609 \pm 93	609 \pm 78
log(TR)m	6.39 \pm 0.12	6.42 \pm 0.13	6.38 \pm 0.13	6.38 \pm 0.12

3.5.1. Analyse de la proportion de réponses correctes

Les moyennes et écart-types des proportions de réponses correctes pour chaque condition expérimentale sont reportés dans le Tableau 2.4. L'analyse ne montre pas d'effet principal de la Congruence sémantique ($\beta = -0.09$, $z = -0.69$, $p = .488$), mais cette variable interagit avec la Similarité physique ($\beta = -0.68$, $z = -2.40$, $p = .017$; Figure 2.13). Plus précisément, les modèles des effets simples montrent qu'il y a un effet d'interférence dans la condition de similarité physique ($\beta = -0.35$, $OR = 1.43$, $z = -2.03$, $p = .042$) mais pas dans la condition de dissimilarité physique ($\beta = 0.32$, $z = 1.47$, $p = .143$).

3.5.2. Analyse des temps de réponses correctes

Les moyennes et écart-types des TR moyens et des $\log(TR)$ moyens pour chaque condition expérimentale sont reportés dans le Tableau 2.4. Le modèle montre un effet principal significatif de la Congruence sémantique (congruent = 623 ± 85 , incongruent = 612 ± 85 ; $\beta = 0.02$, $\%change = 1.80\%$, $t(4419) = 2.65$, $p = .008$). L'effet d'interaction entre la Congruence sémantique et la Similarité physique n'est pas significatif ($\beta = 0.03$, $t(4417) = 1.93$, $p = .054$). Comme l'effet d'interaction n'est pas significatif, nous avons testé l'effet simple de la Congruence sémantique dans chaque condition de Similarité physique en utilisant une correction de Bonferroni du seuil alpha ($0.05/2 = 0.025$). L'effet d'interférence est significatif dans la condition physiquement similaire ($\beta = 0.03$, $\%change = 3.22\%$, $t(2176) = 3.26$, $p = .001$; Figure 2.13), mais pas dans la condition physiquement dissimilaire ($\beta = 0.005$, $t(2227) = -0.01$, $p = .582$).

3.5.3. Analyse des données IRMf

L'objectif principal de ces analyses était d'identifier les régions cérébrales associées à l'effet d'interférence en utilisant le contraste [incongruent > congruent]. Par convention et dans un objectif exploratoire, nous avons également testé le contraste inverse [congruent > incongruent]. Dans la logique des analyses des résultats comportementaux, nous avons examiné ensuite ces deux contrastes dans chaque condition de Similarité physique. Pour chaque cluster d'activation significatif, nous reporterons les coordonnées du pic d'activation du cluster dans l'espace MNI (x, y, z), l'aire de Brodmann correspondante (AB), l'étendue spatiale de l'activation ($k =$ nombre de voxels dans le cluster) et la valeur de t du contraste statistique.

Le contraste [incongruent > congruent] ne révèle aucun cluster d'activation significatif. Par contre, le contraste inverse [congruent > incongruent] montre une activation dans les aires visuelles de l'hémisphère droit. Plus précisément, le gyrus lingual ($12x -73y -10z$, AB 18, $k = 143$, $t = 7.09$; Figure 2.14) est plus activé lorsque les scènes centrale et périphérique sont congruentes que lorsqu'elles sont incongruentes. On retrouve un pattern de résultats semblable lorsqu'on considère la condition expérimentale où les deux scènes sont physiquement dissimilaires. En effet, le contraste [incongruent > congruent]

3. Expériences 2 et 3. Influence d'une scène en vision périphérique sur la catégorisation de scènes en vision centrale

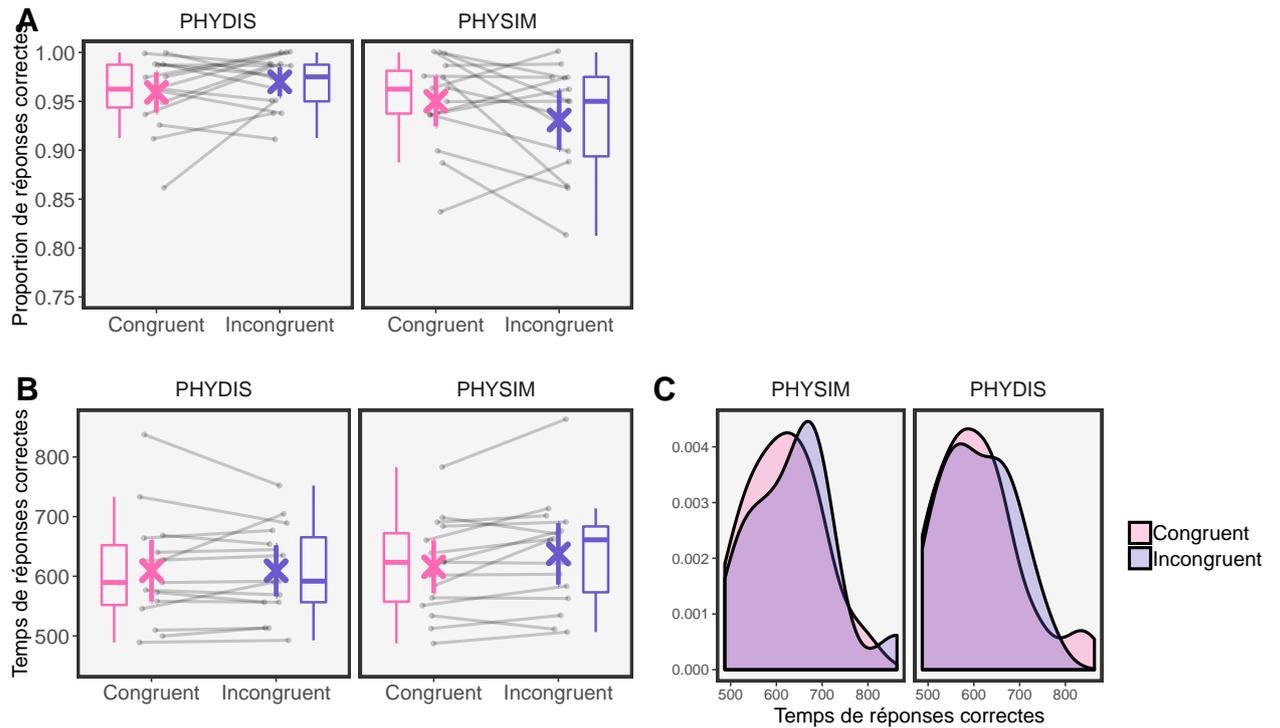


Figure 2.13: Résultats de l'Expérience 3. **A.** Proportion de réponses correctes et **B.** Temps de réponses correctes en fonction de la Congruence sémantique (congruent ; incongruent) et de la Similarité physique (similarité physique, PHYSIM ; dissimilarité physique, PHYDIS) entre les deux scènes. Les \times représentent les moyennes et les barres d'erreur qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **C.** Distributions comparées des temps de réponses correctes pour les conditions congruente et incongruente.

ne révèle aucun cluster significatif. Par contre, le contraste inverse [congruent > incongruent] montre une activation plus étendue dans les aires visuelles de l'hémisphère droit. On observe dans ce cas une activation au niveau de la jonction entre le gyrus fusiforme et le gyrus temporal inférieur (48x -55y -13z, AB 19/37, $k = 52$, $t = 5.26$) et du cuneus (cortex visuel primaire, 12x -91y 11z, AB 47/13, $k = 42$, $t = 4.80$).

Par contre, dans la condition expérimentale où les deux scènes sont physiquement similaires, on observe des activations cérébrales qui rendraient compte de l'effet d'interférence de la vision périphérique observé au niveau comportemental. Dans cette condition, le contraste [incongruent > congruent] montre une activation frontale bilatérale. Ces activations sont localisées dans la partie orbitofrontale du gyrus frontal inférieur (*orbitaris* mais aussi *pars triangularis*) et s'étend jusqu'au cortex insulaire antérieur (hémisphère droit : 33x 11y, 2z, AB 47/13, $k = 131$, $t = 7.59$; hémisphère gauche : -33x 23y 2z, AB 47/13, $k = 58$, $t = 4.11$; Figure 2.14). En revanche, le contraste inverse [congruent > incongruent] ne révèle aucun cluster d'activation significatif.

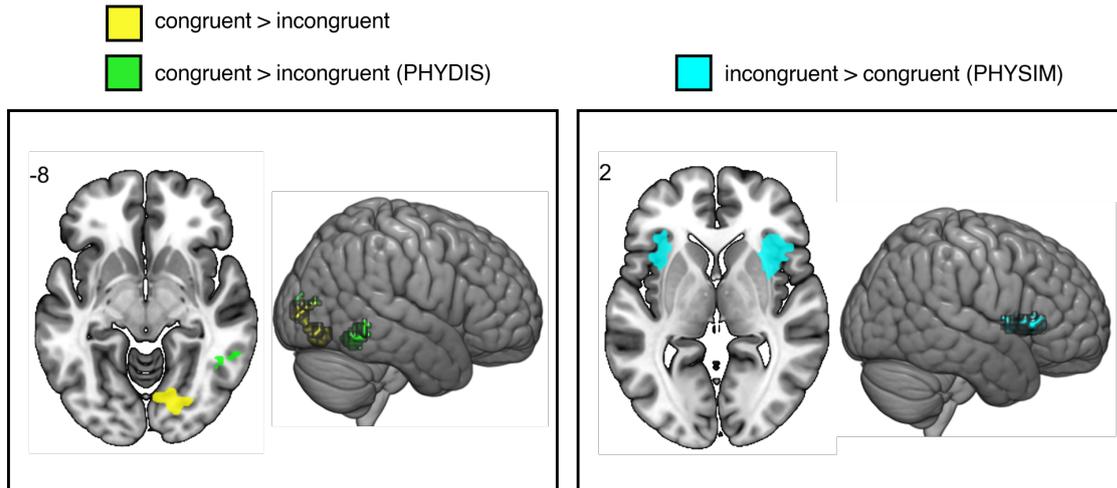


Figure 2.14: Analyse de groupe des données d'IRMf de l'Expérience 3. Régions cérébrales spécifiquement activées pour (1) à gauche, la condition de congruence sémantique [congruent > incongruent] de manière globale et dans la condition de dissimilarité physique (PHYDIS), et (2) à droite pour la condition d'incongruence sémantique (effet d'interférence, [incongruent > congruent] dans la condition de similarité physique (PHYSIM). Les cartes statistiques ont été générées avec un seuil de $p < .05$ avec une correction *Family Wise Error rate* au niveau du cluster et une étendue spatiale d'au moins 10 voxels par cluster.

3.6. Discussion des Expériences 2 et 3

Avec les Expériences 2 et 3, nous avons pour objectif d'étudier si une information rudimentaire disponible en vision périphérique est susceptible d'influencer les traitements réalisés en vision centrale. Les participants catégorisaient une scène cible en vision centrale tandis qu'une scène distractive était présentée en vision périphérique. Nous avons manipulé la congruence sémantique et la similarité physique entre les deux scènes. Notre hypothèse était que l'extraction rapide des BFs en vision périphérique alimente le système de gestion des prédictions. Lorsque cette information est incongruente avec la scène à catégoriser en vision centrale, le signal prédictif renvoyé sur le système de reconnaissance perturberait ainsi la catégorisation (effet d'interférence).

Influence périphérique et effet de la similarité physique

Les résultats comportementaux de ces deux expériences sont cohérents et permettent de tirer les mêmes conclusions. En premier lieu, et conformément à notre hypothèse, nous avons observé un effet d'interférence de la vision périphérique (effet de la congruence sémantique entre la scène cible en vision centrale et la scène distractive en vision périphérique). Dans les deux expériences, les participants font peu d'erreurs (proportion moyenne de réponse correcte autour de 0.95). L'effet d'interférence global sur la proportion de réponses correctes n'a été détecté que dans l'Expérience 2, avec un plus grand échantillon. L'effet d'interférence se manifeste alors surtout sur les temps de réponse.

L'effet d'interférence suggère que même si la scène présentée en vision périphérique est inutile pour réaliser la tâche, son traitement est automatique et influence la catégorisation de scènes en vision centrale. Ces résultats peuvent être expliqués par un modèle FP-peri dans lequel les informations en provenance du champ visuel périphérique seraient automatiquement intégrées aux traitements réalisés en vision centrale par le biais de mécanismes de prédictions.

Ensuite, l'effet d'interférence de la vision périphérique est, dans les deux expériences, dépendant de la similarité physique entre les deux scènes. En effet, nous avons observé un effet d'interférence significatif lorsque les deux scènes étaient physiquement similaires, mais pas lorsqu'elles étaient physiquement dissimilaires. Ainsi, ce résultat indique que les caractéristiques physiques du stimulus périphérique sont susceptibles de moduler le traitement réalisé en vision centrale. Par ailleurs, l'effet d'interaction suggère que ce sont ces propriétés de bas niveau qui seraient utilisées en priorité par le système de gestion des prédictions, puisque les scènes distractives périphériques physiquement dissimilaires n'influençaient pas la catégorisation en vision centrale. Néanmoins, l'intégration de BFs périphériques physiquement dissimilaires (donc d'une information de plus haut niveau) reste plausible. En effet, on peut faire l'hypothèse que des prédictions contenant l'information de plus haut niveau véhiculée par ces scènes soient tout de même utiles, mais que leur effet soit minime et donc que l'absence d'effet dans la condition physiquement dissimilaire soit liée à un problème de design expérimental ou de puissance statistique. Quoi qu'il en soit, l'effet d'interaction avec la similarité physique observé à la fois dans l'Expérience 2 et dans l'Expérience 3 suggère que les représentations issues de l'analyse de BFs périphériques qui influencent la reconnaissance en vision centrale ne seraient pas de nature purement sémantique. Elles conserveraient au contraire certaines propriétés de plus bas niveau du stimulus comme la configuration spatiale des scènes et/ou leurs propriétés spectrales.

Bases cérébrales de l'influence périphérique

Dans l'Expérience 3, nous avons utilisé l'IRM fonctionnelle afin d'étudier les régions cérébrales impliquées dans l'effet d'interférence de la vision périphérique. Les modèles neurocognitifs de la reconnaissance visuelle (Bar 2003; Bar et al. 2006; Kauffmann et al. 2015b, 2014; Peyrin et al. 2010) proposent que les prédictions sont générées grâce au traitement rapide des BFs dans le cortex frontal inférieur, et notamment dans sa partie orbitofrontale. Elles seraient ensuite transmises vers le cortex occipito-temporal où elles seraient comparées au signal ascendant issu du traitement des HFs. Le signal HFs ne correspondant pas aux prédictions (l'erreur de prédiction) serait ensuite transmis vers le cortex frontal inférieur afin de mettre à jour les prédictions. Plus l'erreur de prédiction est grande, et plus les échanges entre ces deux régions devraient être importants, ce qui se manifesterait par une augmentation de l'activité cérébrale.

L'hypothèse d'un rôle prédictif de l'activité précoce du cortex orbitofrontal dans la reconnaissance visuelle est soutenue par l'étude de Bar et al. (2006). Comme nous

l'avons vu dans le Chapitre 1, dans la première expérience de cette étude qui combinait mesures d'IRMf et mesures de MEG, les participants devaient reconnaître des dessins d'objets, présentés rapidement (26 ms) et suivis d'un masque. Les objets étaient difficilement identifiables la première fois qu'ils étaient perçus, mais la reconnaissance devenait possible avec la répétition des stimuli. Les auteurs comparaient ensuite l'activité cérébrale induite par les essais pour lesquels les objets avaient été reconnus à celle induite par les essais pour lesquels ils n'avaient pas été reconnus, et avaient alors observé une activité dans le cortex orbitofrontal qui précédait l'activité du cortex occipito-temporal d'environ 50 ms. Les auteurs en ont conclu que le cortex orbitofrontal permet de rapidement former des prédictions qui facilitent ensuite les processus de reconnaissance d'objet dans le cortex occipito-temporal. C'est aussi dans ce contexte théorique que [Kauffmann et al. \(2015c\)](#) ont étudié à l'aide de l'IRMf les bases cérébrales de l'effet d'interférence. Pour mémoire, les participants de cette étude devaient catégoriser les HFs de scènes hybrides en ignorant le composant BF. Afin de mettre en évidence les régions impliquées dans l'effet d'interférence des BFs, ces auteurs ont soustrait l'activité induite par la condition de congruence sémantique à celle induite par la condition d'incongruence sémantique entre les composants BFs et HFs de l'hybride. L'effet d'interférence des BFs sur la catégorisation des HFs était associé à une activation bilatérale des gyri fusiforme et parahippocampique, du gyrus frontal inférieur, et du cortex orbitofrontal.

Dans la présente étude en IRMf, nous nous attendions à ce que les erreurs de prédiction induites par la scène périphérique s'accompagnent d'une plus forte activation dans le cortex frontal inférieur et le cortex occipito-temporal. Nous avons soustrait l'activité induite par la condition de congruence sémantique à celle induite par la condition d'incongruence sémantique afin de tester l'activation de ces régions. Nos résultats montrent que l'effet d'interférence s'accompagne d'une activation spécifique du cortex frontal inférieur dans les deux hémisphères cérébraux. Cependant, cette activation n'est observée que dans la seule condition de similarité physique entre les deux scènes. Cette observation est le corrélât cérébral des résultats comportementaux qui montrent un effet d'interférence périphérique seulement lorsque les scènes centrales et périphériques sont physiquement similaires. Cette activation bilatérale du cortex frontal inférieur est localisée dans le cortex orbitofrontal et s'étend jusqu'au cortex insulaire antérieur. Les coordonnées des pics d'activation (33x, 11y, 2z dans l'hémisphère droit et -33x 23y 2z dans l'hémisphère gauche) sont proches de celles des pics d'activation observés par [Kauffmann et al. \(2015c\)](#) dans l'hémisphère droit (deux pics d'activations aux coordonnées 29x 25y -36z et 37x 32y 2z). La similitude de ces deux activations pourrait indiquer qu'en accord avec l'hypothèse générale que nous défendons dans cette thèse, il y aurait des mécanismes communs de prédictions basées sur l'analyse rapide des BFs, que celles-ci proviennent du champ visuel central ([Kauffmann et al. 2015c](#)) ou du champ visuel périphérique (Expérience 3). À noter que l'activation orbitofrontale que nous observons est également proche de celle observée par [Bar et al. \(2006 ; -36x 23y -14z\)](#).

Dans l'ensemble, l'activation du cortex orbitofrontal dans notre expérience est donc

cohérente avec celles observées dans des études précédentes visant à étudier l'implication de cette région dans le codage prédictif de l'information visuelle (Bar et al. 2006; Kauffmann et al. 2015c). Il est important de souligner que la résolution temporelle de l'IRMf est faible et ne permet donc pas de savoir à quel moment du processus de catégorisation le cortex orbitofrontal est impliqué. Une interprétation de l'activité du cortex orbitofrontal en termes de mécanismes prédictifs n'est donc valable que dans le cadre de notre hypothèse qui s'appuie sur le modèle FP-peri. Celle-ci a été formulée a priori et elle repose sur un ensemble de résultats de recherche cohérents qui permettent d'envisager le cortex orbitofrontal comme étant une région importante pour la reconnaissance visuelle (voir aussi Chapitre 1 partie 2.3.2). Dans cette mesure, notre interprétation est que l'information de faible résolution spatiale extraite en vision périphérique a pu être rapidement projetée vers le cortex orbitofrontal pour former des prédictions qui étaient pertinentes (condition de congruence sémantique) ou non pertinentes (condition d'incongruence sémantique). Dans le cas d'une prédiction non pertinente, le signal prédictif comparé aux informations ascendantes issues des HFs de la scène centrale dans le cortex occipito-temporal mènerait à une erreur de prédiction et à des échanges plus importants entre ces régions et le cortex orbitofrontal. Par contre, à l'inverse des précédentes études réalisées dans ce contexte de recherche, nous n'avons pas observé d'activité dans le cortex occipito-temporal. L'utilisation d'un seuil de significativité plus libéral ($p < .005$, non corrigé) dévoile néanmoins un cluster d'activation dans le gyrus fusiforme gauche ($-36x -52y -16z$, AB 37, $k = 22$, $t = 3.97$). Il est possible que l'absence d'activation significative dans ces régions puisse être liée à la facilité de la tâche comportementale, qui donne lieu à des proportions moyennes de réponses correctes d'environ 0.95.

Pour conclure fortement sur le rôle du cortex orbitofrontal, il reste toutefois nécessaire d'accéder à une meilleure résolution temporelle, en couplant par exemple le paradigme d'interférence avec une technique comme l'EEG, comme l'avaient fait Mu et Li (2013). Une telle étude permettrait de mieux désambiguïser l'hypothèse d'un rôle prédictif du cortex orbitofrontal d'explications alternatives, comme par exemple l'hypothèse d'un rôle inhibiteur de cette région. En effet, il est possible que des processus d'inhibition interviennent après la reconnaissance, à un niveau décisionnel, afin de contrôler la réponse motrice.

Additionnellement, nos analyses ont révélé que la condition de congruence sémantique (par rapport à la condition d'incongruence sémantique) était accompagnée d'activations postérieures dans l'hémisphère droit dans le cas spécifique où les scènes centrale et périphériques étaient physiquement dissimilaires. Ces activations comprenaient des clusters relativement petits localisés dans le gyrus fusiforme et le gyrus temporal inférieur ainsi que dans la scissure calcarine, qui correspond approximativement au cortex visuel primaire. Dans cette condition, il y a la présence de deux scènes de même catégorie sur-ordonnée (e.g., deux scènes naturelles) mais dont l'arrangement spatial et les propriétés spectrales ne se ressemblent pas (e.g., une scène de forêt, avec beaucoup d'énergie dans les HFs et dans les orientations verticales *vs* une scène de plage, avec de grands

blobs uniformes en BFs). La présence de propriétés visuelles variées ne serait alors pas compatible avec l'unique information abstraite de haut niveau (catégorie 'naturelle') et pourrait alors provoquer un surinvestissement des régions visuelles.

Conclusion

Pour conclure, ces deux expériences nous permettent de soutenir le modèle FP-peri en montrant que l'information en vision périphérique est traitée automatiquement lors de la catégorisation de scènes en vision centrale et influence cette catégorisation. Par ailleurs, cette influence semble reposer sur des informations visuelles de bas niveau liées aux caractéristiques physiques des scènes, comme le suggère le plus fort effet d'interférence dans la condition de similarité physique. L'Expérience 3 permet d'appuyer l'hypothèse d'un substrat neuronal au système de gestion des prédictions du modèle FP dans le cortex orbitofrontal, cette région étant plus fortement activée en cas d'incongruence sémantique.

Il faut souligner que le paradigme expérimental que nous avons utilisé, bien qu'il permette de contrôler assez précisément certains paramètres comme la similarité physique des scènes, est assez peu écologique. Par exemple, nous n'avons pas tenu compte du phénomène de magnification corticale. En effet, les projections des informations visuelles sont déformées sur le cortex visuel. Une grande surface du cortex visuel est consacrée au traitement de l'information disponible en vision centrale tandis qu'une plus petite surface est dédiée aux informations disponibles en vision périphérique. Par conséquent, pour stimuler le même nombre de cellules du cortex visuel primaire, la taille des scènes présentées en vision périphérique doit être plus grande que celle des scènes présentées en vision centrale. Ce n'était pas le cas dans les Expériences 2 et 3. Il est alors possible que nous ayons sous-estimé la taille de l'effet d'interférence, et que l'influence de la vision périphérique soit encore plus importante lorsque l'information disponible couvre une plus grande part du champ visuel périphérique. Dans la section suivante, nous présentons deux expériences dans lesquelles nous avons utilisé des stimuli plus écologiques, qui couvrent une plus grande partie du champ visuel périphérique.

4. Expériences 4 et 5 : Influence de l'information contextuelle d'une scène en vision périphérique sur la catégorisation d'un objet en vision centrale

N.B. Dans cette partie expérimentale, nous reportons les données de deux expériences publiées dans la revue *Journal of Vision* (Roux-Sibilon et al. 2019 ; Annexe A). Dans cet article, nous avons testé nos hypothèses à l'aide d'analyses de la variance à mesures répétées. À l'occasion de la rédaction de ce manuscrit de thèse, nous avons réanalysé ces données en utilisant des modèles à effets mixtes (cf. Section 1).

4.1. Objectifs et hypothèses de l'Expérience 4

Cette expérience avait pour objectif d'étudier comment la présence d'informations contextuelles en vision périphérique pouvait influencer la reconnaissance d'objets en vision centrale. Au delà de tester les effets de contexte classiquement étudiés dans le domaine de la reconnaissance visuelle (cf. Chapitre 1 sous-partie 3.2), l'intérêt méthodologique est que l'information de contexte (la scène) était présentée seulement en vision périphérique (i.e. au delà des six ou huit degrés d'angle visuel centraux), et que la taille rétinienne de l'objet était relativement petite (environ trois degrés d'angle visuel) et celui-ci présenté seulement en vision centrale. Ces paramètres expérimentaux permettent de contrôler les biais discutés dans la partie 4.1 du Chapitre 1 (cf. aussi 1.1) qui limitent l'inscription des effets de contexte dans le cadre théorique d'un modèle FP-peri. Cela permet également de proposer une situation expérimentale plus écologique que celle des Expériences 2 et 3, puisque les stimuli couvrent une plus grande partie du champ visuel périphérique et permettent donc de mieux imiter une situation naturelle où une grande quantité d'information est disponible en vision périphérique alors que la vision centrale est dédiée à l'analyse d'une plus petite quantité d'information (ici, un objet). Le paradigme expérimental est inspiré de celui des expériences précédentes de ce chapitre, et des travaux portant sur les effets de contexte (e.g., Boucart et al. 2013; Davenport 2007; Davenport & Potter 2004). Nous avons demandé à des participants de catégoriser un objet apparaissant dans leur vision centrale alors qu'une scène sémantiquement reliée (congruente/prédictive) ou non-reliée (incongruente/non-prédictive) apparaissait en vision périphérique. Les participants devaient catégoriser l'objet en vision centrale comme étant un meuble ou un animal (Figure 2.15). Nous avons mesuré l'effet d'interférence, correspondant à de meilleures performances lorsque les informations contextuelles en vision périphérique sont congruentes avec l'objet (e.g., une vache dans un champ)

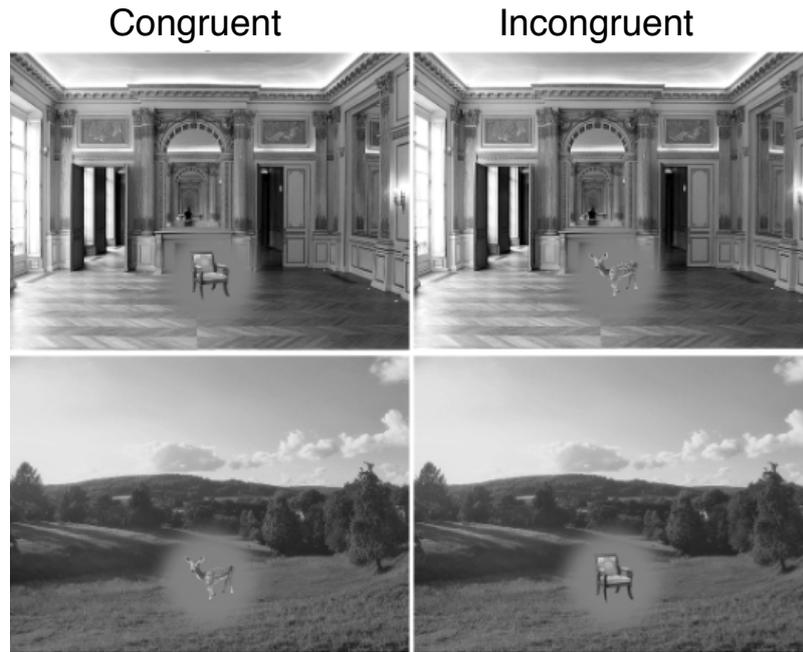


Figure 2.15: Stimuli l'Expérience 4. Exemple d'un groupe de quatre associations scène-objet. Les associations pouvaient être congruentes (une scène d'intérieur et un meuble / une scène d'extérieur et un animal) ou incongruentes (une scène d'intérieur et un animal / une scène d'extérieur et un meuble)

que lorsqu'elles sont incongruentes (e.g., une vache dans un salon). Comme dans les Expériences 2 et 3, les scènes n'étaient pas filtrées, puisque nous avons considéré que la présentation des scènes en vision périphérique agirait comme un filtre passe-bas naturel. Notre hypothèse était qu'une représentation *gist* formée rapidement et automatiquement sur la base du traitement des BFs de la scène en vision périphérique permettrait de générer des prédictions qui viendraient ensuite faciliter le traitement de l'objet en vision centrale. Nous nous attendions donc à observer un effet d'interférence.

Un autre objectif de l'Expérience 4 était de tester comment le temps pendant lequel la scène était traitée avant l'apparition de l'objet pouvait moduler l'effet d'interférence, en manipulant le SOA (*Stimulus Onset Asynchrony*) entre la scène et l'objet. La scène était ainsi présentée 150 ms avant l'apparition de l'objet (condition SOA-150ms), 30 ms avant l'apparition de l'objet (condition SOA-30ms) ou de manière simultanée (condition SOA-0ms). Nous nous attendions à ce que si l'information en vision périphérique est utilisée pour générer un signal prédictif qui module le traitement visuel de l'objet en vision centrale, l'influence périphérique sur la catégorisation de l'objet (i.e. l'effet de congruence) serait plus grande lorsque l'information périphérique peut être accumulée pendant plus de temps - i.e. lorsque la scène est présentée plus longtemps avant l'apparition de l'objet - permettant ainsi une représentation plus robuste du contexte. Nous avons donc testé l'effet d'interaction entre la congruence sémantique et le SOA.

4.2. Expérience 4 : Méthode

4.2.1. Participants

Cinquante participants (43 femmes ; âge moyen \pm écart-type : 21 ± 4 ans ; 18-26 ans) ont participé à l'expérience. Nous avons fixé la taille d'échantillon en nous basant sur la taille de l'effet d'interférence que nous avons observé dans une expérience pilote avec les mêmes stimuli. Nous avons ainsi réalisé une analyse de puissance avec une taille d'effet attendue de 0.62 (d de Cohen pour effets intra-sujets), une puissance désirée de 0.99, et un seuil alpha de 0.05. Tous les participants étaient droitiers et avaient une vision normale ou corrigée à la normale et pas d'antécédents de maladies neurologiques ou psychiatriques. Avant de participer à l'expérience, chaque participant a donné son consentement libre et éclairé en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine. À la fin de l'expérience, chaque participant a reçu un bon d'expérience de 1 point à faire valoir lors des examens universitaires.

4.2.2. Stimuli

Les stimuli étaient des images composées d'un objet et d'une scène en arrière-plan. Toutes les images utilisées pour créer les stimuli (scènes d'une part et objets d'autre part) ont été téléchargées à partir de la banque d'image Pixabay (<https://pixabay.com/fr/>) sous licence Creative Commons Zero, ou à partir de recherches Google d'images libres de droits. Les scènes utilisées comme arrière-plans étaient des photographies représentant des paysages naturels d'extérieurs (e.g., savane, champ, banquise) ou des vues d'intérieurs (e.g., salon, couloir). Nous les avons converti en 256 niveaux de gris et redimensionné à 1024×768 pixels. Les stimuli étaient présentés sur un écran 30 pouces (Dell UltraSharp) avec une résolution de 2500×1600 et un taux de rafraîchissement de 60 Hz. Nous avons estimé la fonction gamma de l'écran en mesurant avec un outil de calibration (Spyder5Elite) les valeurs de luminance de l'affichage uniforme de différents niveaux de gris générés dans MATLAB. En nous basant sur cette fonction, nous avons corrigé les valeurs de luminance de chaque scène pour linéariser indirectement la fonction gamma de l'écran¹. La luminance moyenne et le contraste RMS des scènes étaient ensuite égalisés pour obtenir une luminance moyenne de 0.51 pour des valeurs de luminance allant de 0 à 1 (i.e. 130 sur une échelle de 256 niveaux de gris) et un RMS moyen de 0.24 (i.e. 61 sur une échelle de 256 niveaux de gris). Ces valeurs correspondaient aux valeurs moyennes de luminance et de RMS de la base entière d'images de scènes (paysages et vues d'intérieurs). Les objets étaient des animaux (e.g., vache, marmotte) et des meubles (e.g., fauteuil, table basse) qui étaient détournés et extraits des photographies en utilisant Adobe Photoshop. Chaque objet était redimensionné afin que son côté le plus large couvre 128 pixels (soit la largeur, soit la hauteur, selon l'objet).

1. À noter que la calibration de l'écran n'a été réalisée qu'à partir de l'Expérience 4, le LPNC ne disposant pas de l'outil de calibration avant.

Nous avons créé 20 groupes de quatre stimuli "scène-objet" (80 stimuli) en combinant une scène avec un objet. Les quatre combinaisons donnaient deux associations scène-objet congruentes (animal + extérieur et meuble + intérieur) et deux associations incongruentes (animal + intérieur et meuble + extérieur; Figure 2.15-A). Les scènes et les objets qui formaient des associations congruentes étaient associés sur la base de leur contingence dans la vie réelle (e.g., une vache et un champ, un pingouin et une banquise). Chaque objet était collé sur une scène congruente et une scène incongruente, au centre sur l'axe horizontal, mais dans la partie basse de l'image sur l'axe vertical, afin de respecter la position typique des objets sur les photographies. L'objet était isolé de la scène par un patch gris circulaire (d'un diamètre d'environ 250 pixels) dont les bords étaient progressivement intégrés dans la scène (Figure 2.15-A). Le fait d'isoler l'objet nous a permis de restreindre l'information contextuelle à la seule vision périphérique, et de contrôler les potentiels effets de contraste local qui auraient pu améliorer la saillance de l'objet dans une condition plus que dans l'autre.

4.2.3. Procédure

L'expérience était programmée avec E-Prime 2. Les participants voyaient les stimuli à une distance de 50 cm. À cette distance, les scènes en arrière-plan couvraient 24° d'angle visuel horizontalement et 18° verticalement. Le côté le plus large de l'objet (soit la largeur soit la hauteur) couvrait 3° et le patch circulaire gris qui isolait l'objet de la scène couvrait environ 6° en diamètre. Pour maintenir la distance à l'écran, la tête des participants reposait sur une mentonnière positionnée dans l'axe sagittal médian.

Les participants avaient pour consigne de fixer une petite croix blanche, centrée sur la position où l'objet allait apparaître (i.e. au centre de l'écran horizontalement, et dans la partie basse verticalement). Chaque essai commençait par l'apparition de cette croix de fixation présentée pendant 500 ms, suivie de la séquence de stimuli et d'un masque de la taille de la scène pendant 30 ms afin d'éviter la persistance rétinienne. Le masque était constitué d'un bruit distribué en $1/f$ afin d'imiter la distribution du contraste en fonction de la fréquence spatiale dans les scènes naturelles (Ruderman & Bialek 1994). L'intervalle interstimulus était de 2500 ms. La séquence de stimuli dépendait de la condition expérimentale. Dans la condition SOA-0ms, le stimulus scène-objet était présenté pendant 150 ms. Dans la condition SOA-30ms, le stimulus scène-objet était également présenté pendant 150 ms, mais était précédé par la scène seule (incluant le patch circulaire gris) pendant 30 ms. Similairement, dans la condition SOA-150ms, le stimulus scène-objet était présenté pendant 150 ms, et était précédé par la scène seule (avec le patch circulaire gris) pendant 150 ms (Figure 2.16). Les trois conditions de SOA étaient manipulées en blocs, et l'ordre des blocs était contrebalancé à travers les participants par une procédure en carré latin. Les participants devaient catégoriser l'objet (animal *vs* meuble) en appuyant sur deux boutons d'un clavier d'ordinateur. Ils n'avaient pas de consigne particulière au sujet de la scène en arrière-plan, mais s'ils posaient la question, on leur disait que celle-ci n'avait pas d'importance dans la tâche. Les boutons

de réponse étaient contrebalancés à travers les participants. L'exactitude de la réponse et le temps de réponse était enregistrés à chaque essai. Il y avait 240 essais au total (2 conditions de Congruence \times 3 conditions de SOA \times 20 groupes de stimuli \times 2 catégories d'objets), soit 40 par condition expérimentale. Les participants pouvaient prendre des pauses entre chaque bloc de SOA s'ils le désiraient, et l'expérience durait environ 20 minutes. Avant de commencer, les participants réalisait une session d'entraînement (12 essais) afin de se familiariser à la tâche et aux stimuli.

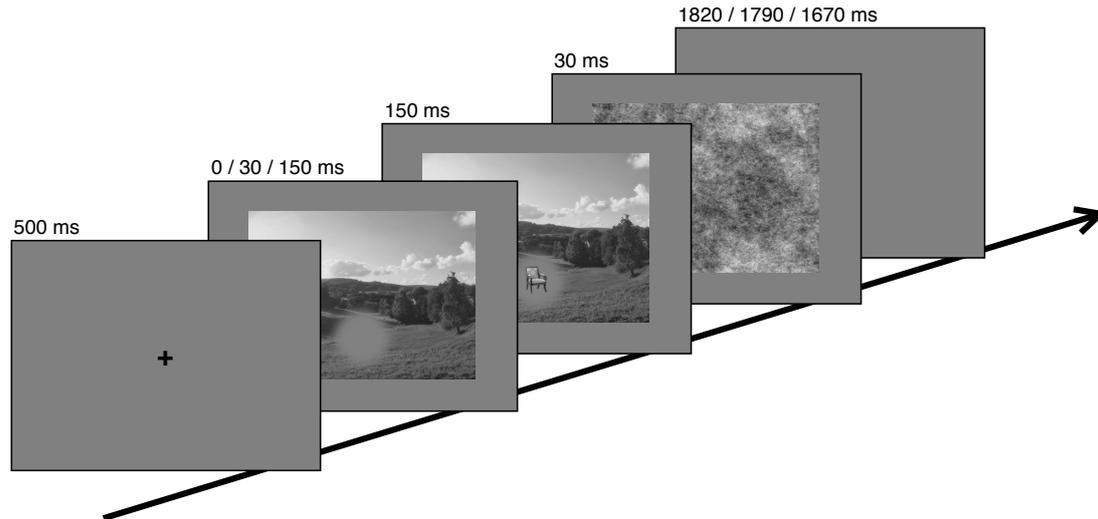


Figure 2.16: Procédure expérimentale de l'Expérience 4.

4.3. Expérience 4 : Résultats

Les effets fixes des modèles mixtes sont estimés pour la Congruence sémantique entre l'objet et la scène (congruent, incongruent), le SOA (0ms, 30ms, 100ms), ainsi que l'interaction entre ces deux facteurs, et les effets aléatoires sont pris en compte pour l'ordonnée à l'origine des participants.

Tableau 2.5: Résultats de l'Expérience 4. Moyennes \pm écart-types de la proportion de réponses correctes (pRC), des temps de réponses correctes moyens en millisecondes (TRm), et des TR transformés (transformation logarithmique) moyens (log(TR)m) en fonction de la Congruence sémantique (congruent ; incongruent) et du SOA (0 ms ; 30 ms ; 150 ms).

SOA :	0 ms	0 ms	30 ms	30 ms	150 ms	150 ms
	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
pRC	0.96 \pm 0.20	0.96 \pm 0.20	0.97 \pm 0.18	0.96 \pm 0.20	0.96 \pm 0.19	0.95 \pm 0.23
TRm (ms)	567 \pm 159	568 \pm 158	539 \pm 149	543 \pm 140	527 \pm 135	547 \pm 130
log(TR)m	2.74 \pm 0.10	2.74 \pm 0.10	2.72 \pm 0.10	2.72 \pm 0.10	2.71 \pm 0.10	2.73 \pm 0.09

4.3.1. Analyse de la proportion de réponses correctes

Les moyennes et écart-types du taux de réponses correctes pour chaque condition expérimentale sont reportés dans le Tableau 2.5. L'effet principal de la Congruence sémantique est significatif, la proportion de réponses correctes est légèrement plus grande dans la condition congruente que dans la condition incongruente (congruent = 0.96 ± 0.19 , incongruent = 0.955 ± 0.21 ; $\beta = 0.26$, OR = 1.30, $z = 2.75$, $p = .006$). Le SOA n'interagit pas avec la Congruence ($\beta = 0.39$, $z = 1.71$, $p = .088$). En général, la proportion de réponses correctes était élevée (en moyenne 0.96 ± 0.20), avec des distributions reflétant un effet plafond (cf. Figure 2.17-A)

4.3.2. Analyse des temps de réponses correctes

Les moyennes et écart-types des TRm et des log(TR)m pour chaque condition expérimentale sont reportés dans le Tableau 2.5. Suite à la procédure d'écrémage, 1.76% des temps de réponse ont été retirés des données utilisées pour les analyses statistiques. L'effet principal de la Congruence sémantique est significatif (congruent = 544 ± 149 ms, incongruent = 553 ± 144 ms; $\beta = 0.009$, %change = 0.87%, $t(11260) = 6.31$, $p < .001$). Par ailleurs, le modèle d'interaction montre que le SOA interagit avec la Congruence ($\beta = 0.02$, $t(11260) = 4.90$, $p < .001$; Figure 2.17-B et -C). Nous avons utilisé un modèle linéaire et un modèle quadratique pour tester si l'augmentation du SOA faisait augmenter linéairement l'effet d'interférence (i.e. la différence [incongruent - congruent]). Le modèle linéaire montre un effet significatif ($t(49)$, 4.70, $p < .001$) alors que le modèle quadratique n'est pas significatif ($t(49) = 1.24$, $p = 0.223$) indiquant que l'influence de la scène périphérique sur la catégorisation de l'objet augmente linéairement avec le SOA. Les modèles des effets simples montrent qu'il n'y a pas d'effet de la Congruence dans la condition SOA-0ms ($\beta = 0.002$, $t(3714) < 1$). En revanche, l'effet de la Congruence est significatif dans la condition SOA-30ms, bien que de très faible magnitude (congruent = 539 ± 149 ms, incongruent = 543 ± 140 ms; $\beta = 0.006$, %change = 0.61%, $t(3740) = 2.64$, $p = .008$), et il est également significatif dans la condition SOA-150ms, avec cette fois un effet de plus grande magnitude (congruent = 527 ± 135 ms, incongruent = 547 ± 130 ms; $\beta = 0.02$, %change = 1.80%, $t(3740) = 7.97$, $p < .001$).

4.4. Résumé et limites de l'Expérience 4

Pour résumer, les résultats de l'Expérience 4 ont montré que les participants catégorisaient les objets en vision centrale plus rapidement et plus précisément lorsqu'une scène congruente (plutôt qu'incongruente) était traitée en vision périphérique juste avant la présentation de l'objet (conditions SOA-30ms et SOA-150ms). Cet effet suggère que lorsqu'ils catégorisaient les objets, les participants ont traité automatiquement l'information en vision périphérique, malgré qu'elle ne soit pas utile pour réaliser la tâche. Par contre, lorsque la scène apparaissait en même temps que l'objet (condition SOA-0ms), nous n'avons pas détecté d'effet de la congruence. Nous remarquons que les effets sont de faible magnitude (%change = 0.61% pour la condition SOA-30ms et 1.80% pour la

4. Expériences 4 et 5 : Influence de l'information contextuelle d'une scène en vision périphérique sur la catégorisation d'un objet en vision centrale

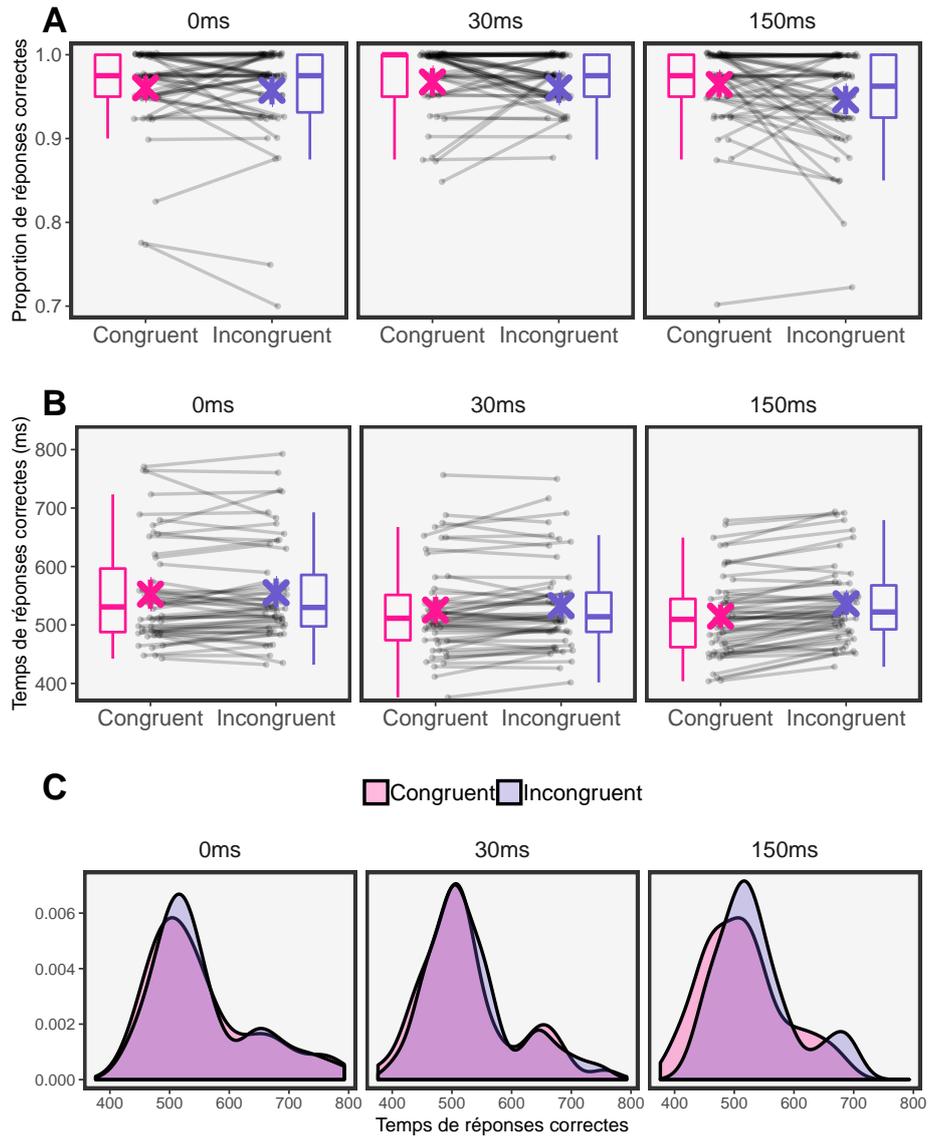


Figure 2.17: Résultats de l'Expérience 4. **A.** Proportion de réponses correctes et **B.** Temps de réponses correctes (ms) en fonction de la Congruence sémantique (congruent ; incongruent) et du SOA (0 ms ; 30 ms ; 150 ms). Les \times représentent les moyennes et les barres verticales qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **C.** Distributions comparées des temps de réponses correctes pour la condition congruente et la condition incongruente, dans les trois conditions de SOA.

condition SOA-150ms). Les résultats de cette expérience indiquent donc que l'information contextuelle en vision périphérique pourrait être utile lors de la catégorisation d'objet en vision centrale lorsqu'elle est perçue légèrement en amont, mais que son impact est relativement limité. Nous reviendrons sur la taille des effets d'interférence périphérique dans le Chapitre 4.

Dans l'Expérience 4, les mécanismes qui sous-tendent l'effet de congruence ne sont

pas clairs. D'abord, nous n'avons pas testé de condition contrôle, sans information contextuelle en vision périphérique, ce qui ne nous permet pas d'avoir des informations sur le sens de l'effet de congruence : la catégorisation d'objets est elle facilitée dans le cas d'un contexte périphérique congruent ou au contraire gênée dans le cas d'un contexte périphérique incongruent ? Par ailleurs, un problème substantiel avec ce paradigme de présentation séquentielle est que nous ne pouvons être sûr des mécanismes responsables de la réponse du participant. En effet, il est possible que dans une certaine proportion d'essais, la catégorisation rapide et implicite de la scène ait pu directement entraîner la réponse (i.e. par simple association des catégories intérieur ↔ meuble et extérieur ↔ animal), sans qu'aucune modulation prédictive sur les processus de reconnaissance de l'objet n'ait été impliquée. Le temps durant lequel la scène est perçue avant l'apparition de l'objet pourrait dans ce cas permettre d'amorcer la réponse, basée seulement sur la catégorie de la scène et pas sur celle de l'objet. Un cas extrême (et hypothétique) de cette stratégie de 'précipitation' serait une situation où un participant répond toujours 'animal' lorsqu'une scène d'extérieur est reconnue et toujours 'meuble' lorsqu'une scène d'intérieur est reconnue, indépendamment de la catégorie de l'objet. Dans ce cas extrême, la proportion de réponses correctes serait de 100% dans la condition congruente et de 0% dans la condition incongruente et en conséquence, l'effet de la congruence sémantique serait maximal. Compte tenu de ce qui précède, l'effet de la congruence que nous avons observé ici n'est peut être pas entièrement dû à l'intégration de l'entrée visuelle périphérique pendant la catégorisation de l'objet en vision centrale. Malheureusement, le design expérimental de cette expérience ne nous permet pas de dissocier les processus de catégorisation d'objet influencés par l'information contextuelle en vision périphérique (réponse intégrée) d'une réponse entraînée par la simple catégorisation de la scène (réponse précipitée). Pour finir, les résultats de l'Expérience 4 montrent un effet plafond sur la proportion de réponses correctes, ce qui suggère que la tâche de catégorisation au niveau de base était trop facile à réaliser, et rend difficile la détection d'effets d'interaction. Ces différentes limites ont été prises en compte dans l'Expérience 5 qui suit.

4.5. Expérience 5 : Objectifs et hypothèses

Cette expérience avait pour objectif d'une part de répliquer l'effet d'interférence contextuelle observé dans l'Expérience 4, et d'autre part de mieux comprendre comment l'information en vision périphérique peut influencer les traitements en vision centrale, en considérant les limites de l'Expérience 4. Nous avons utilisé le même paradigme expérimental. Toutefois, les scènes étaient toujours présentées pendant 150 ms avant l'apparition de l'objet (i.e. la condition dans laquelle nous avons observé un effet de congruence robuste dans l'Expérience 4).

4.5.1. Mesure du seuil de visibilité

Dans l'Expérience 4, l'analyse de la proportion de réponses correctes témoignait d'un effet plafond, indiquant que la tâche que nous utilisons est très facile à réaliser (catégorisation sur-ordonnée meuble/animal). Par ailleurs, les objets étaient non ambigus et de bonne qualité perceptive. Dans ces conditions, il est possible que le poids des influences prédictives soit minimisé par rapport à celui des entrées sensorielles. Le raisonnement est que le poids des prédictions devrait être plus important relativement à celui des entrées sensorielles lorsque l'entrée visuelle est pauvre ou ambiguë (Kok & de Lange 2015). Dans cette idée, des études ont montré que les processus prédictifs améliorent la perception visuelle lorsque le stimulus est bruité ou incomplet (e.g., Brandman & Peelen 2017; Tang et al. 2018; Teufel, Dakin, & Fletcher 2018; Wyatte, Curran, & O'Reilly 2012). L'information en vision périphérique devrait donc avoir une plus forte influence lorsque l'information en vision centrale est de mauvaise qualité. Ainsi, dans l'Expérience 5, nous avons voulu maximiser le poids des influences prédictives. Nous avons diminué la visibilité de l'objet de manière paramétrique en manipulant un paramètre qui permet de diminuer la cohérence de la phase de l'objet. Les objets étaient ainsi présentés à sept niveaux de visibilité différents (Figure 2.18). Cette manipulation paramétrique nous a permis d'estimer des fonctions psychométriques liant la proportion de réponses correctes au niveau de visibilité de l'objet. Nous avons ainsi pu estimer le seuil de visibilité pour atteindre 75% de réponses correctes, et nous avons comparé ce seuil entre les conditions congruente et incongruente (effet de la congruence sur le seuil de visibilité).

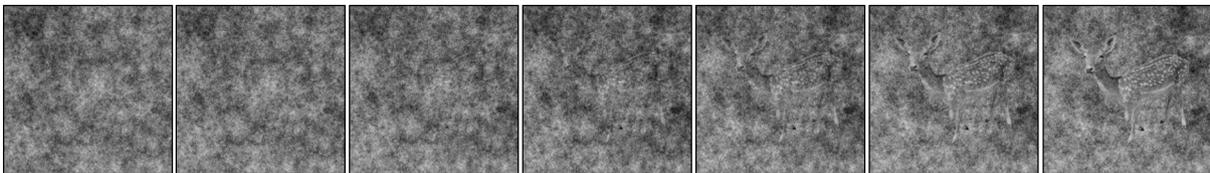


Figure 2.18: Illustration de la modulation paramétrique de la cohérence de phase des objets dans l'Expérience 5. Chaque objet (ici, un faon) était présenté à sept niveaux de visibilité différents.

4.5.2. Direction de l'effet de congruence

Pour pouvoir évaluer la direction de l'effet de congruence (facilitation dans la condition congruente, gêne dans la condition incongruente, ou les deux), nous avons inclus une condition expérimentale permettant d'avoir une ligne de base. Dans cette condition, nous avons combiné l'objet avec un arrière-plan bruité, afin qu'aucune information pertinente ne soit perçue en périphérie. Le bruit était distribué en $1/f$ afin d'imiter la distribution du contraste en fonction de la fréquence spatiale dans les scènes naturelles (Ruderman & Bialek 1994).

4.5.3. Contrôle expérimental : estimation de l'importance des réponses précipitées

Nous avons évoqué précédemment que le paradigme de présentation séquentielle utilisé dans les Expériences 4 et 5 pouvait provoquer une stratégie de réponse précipitée. Afin de distinguer le composant de la mesure dû à la seule catégorisation de la scène périphérique (réponses précipitées) du composant effectivement dû à l'influence de la scène périphérique sur la catégorisation de l'objet (réponses intégrées), nous avons inclus des essais dans lesquels il n'y avait pas d'objet dans l'image. Dans ces essais, la région de l'image où l'objet est censé se trouver était remplie de bruit $1/f$. Dans ce cas, il n'y avait évidemment pas de réponse correcte, mais puisque ces essais étaient inclus aléatoirement parmi les essais dans lesquels un objet était présent dans l'image (mais souvent très peu visible), les participants ne se rendaient pas compte de l'absence d'objet. Nous avons testé si, lors de ces essais pour lesquels il n'y avait pas d'objet dans le stimulus, les participants avaient tendance à se baser sur la catégorisation de la scène pour donner leur réponse de catégorisation, et si cette tendance corrélait avec l'effet de la congruence sémantique. Une corrélation positive suggérerait que l'effet de la congruence est en partie dû au simple traitement de la scène, sans qu'il n'y ait de modulation prédictive pendant le traitement de l'objet (i.e. sans qu'il n'y ait de véritable effet d'interférence pour une certaine proportion d'essais).

4.5.4. Nature de l'influence prédictive

Dans l'Expérience 5, comme dans les Expériences 2 et 3, nous nous sommes aussi intéressés à la nature de l'influence périphérique. Les résultats des Expériences 2 et 3 suggèrent que les prédictions issues de la vision périphérique contiennent des informations sur les propriétés de bas niveau du stimulus visuel. Dans l'Expérience 5, nous avons cette fois voulu étudier si les propriétés visuelles de bas niveau des scènes peuvent suffire à enclencher des prédictions qui aident la catégorisation des objets congruents, ou si au contraire, la possibilité d'extraire une signification / une représentation sémantique est indispensable. Les deux types d'influences sont plausibles. À l'appui de l'hypothèse d'influences issues de propriétés visuelles de bas niveau, les théories prédictives de perception visuelle (Friston & Stephan 2007; T. S. Lee & Mumford 2003; Rao & Ballard 1999) postulent que les aires de bas niveau du cortex visuel sont elles-mêmes susceptibles d'activer des prédictions et de les envoyer sur les aires de plus bas niveau dans la hiérarchie du traitement visuel (e.g., V2 sur V1). À l'appui de l'hypothèse d'influences nécessairement sémantiques, les modèles neurocognitifs de la reconnaissance visuelle (Bar 2003; Bar et al. 2006; Kauffmann et al. 2014; Peyrin et al. 2010), postulent que les prédictions sont plutôt déclenchées dans le cortex frontal inférieur, et ces modèles proposent que cette région coderait plutôt les aspects sémantiques du stimulus.

Afin de décrire les propriétés de bas niveau de nos stimuli, nous avons moyenné les spectres d'amplitude de nos stimulus en les regroupant par catégories (scènes d'extérieur, scènes d'intérieur, meubles, animaux). Comme on peut le voir sur la Figure 2.19, les

images de meubles et de scènes d'intérieur ont tendance à avoir des spectres d'amplitude comparables, avec une énergie concentrée principalement sur les orientations verticales, horizontales mais aussi obliques à 45° (à cause de la perspective), allant des plus basses aux plus hautes fréquences spatiales. Parallèlement, les images d'animaux et de scènes d'extérieur ont tendance à avoir des spectres d'amplitude comparables, avec une énergie distribuée plus équitablement à travers les différentes orientations, et concentrée surtout dans les fréquences spatiales les plus basses. Vu la similarité des spectres d'amplitude de ces catégories de scènes et des objets qui y sont associés, il est possible que les influences de la vision périphérique sur la reconnaissance des objets en vision centrale soient basées seulement sur les propriétés de bas niveaux des stimuli, plutôt que sur une représentation sémantique de la scène.

Dans le but de tester l'influence des deux types d'information (bas niveau *vs* sémantique), nous avons manipulé la présence d'un contenu sémantique dans la scène périphérique. Dans la condition *intacte*, nous avons utilisé les images originales des scènes. Dans la condition *scrambled*, nous avons empêché la possibilité d'extraire des informations sémantiques de l'image en brouillant la phase des images intactes par permutation aléatoire (Figure 2.19). Cette procédure permet de préserver le contenu en orientation et en fréquences spatiales tout en empêchant de traiter une information sémantique (Goffaux et al. 2010; Woodhead, Wise, Sereno, & Leech 2011). Si les propriétés de bas niveau de la scène peuvent suffire à déclencher des prédictions, nous devrions observer un effet d'interférence aussi bien dans la condition *scrambled* et dans la condition *intacte*. Si l'influence périphérique implique au contraire des représentations sémantiques, nous avons prédit un plus grand effet de congruence dans la condition *intacte* que dans la condition *scrambled*.

4.6. Expérience 5 : Méthode

4.6.1. Participants

Dix-huit participants (14 femmes; 21 ± 3 ans) ont participé à l'expérience. La taille d'échantillon a été choisie sur la base d'une analyse de puissance avec une taille d'effet attendue de 1.08 (d de Cohen pour échantillons appariés pour la taille de l'effet de congruence dans la condition SOA-150ms de l'Expérience 4), une puissance désirée de 0.99, et un seuil alpha de 0.05. Tous les participants étaient droitiers et avaient une vision normale ou corrigée à la normale et pas d'antécédents de maladies neurologiques ou psychiatriques. Avant de participer à l'expérience, tous ont donné leur consentement libre et éclairé en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine. À la fin de l'expérience, chaque participant a reçu un bon d'expérience de 1 point à faire valoir lors des examens universitaires.

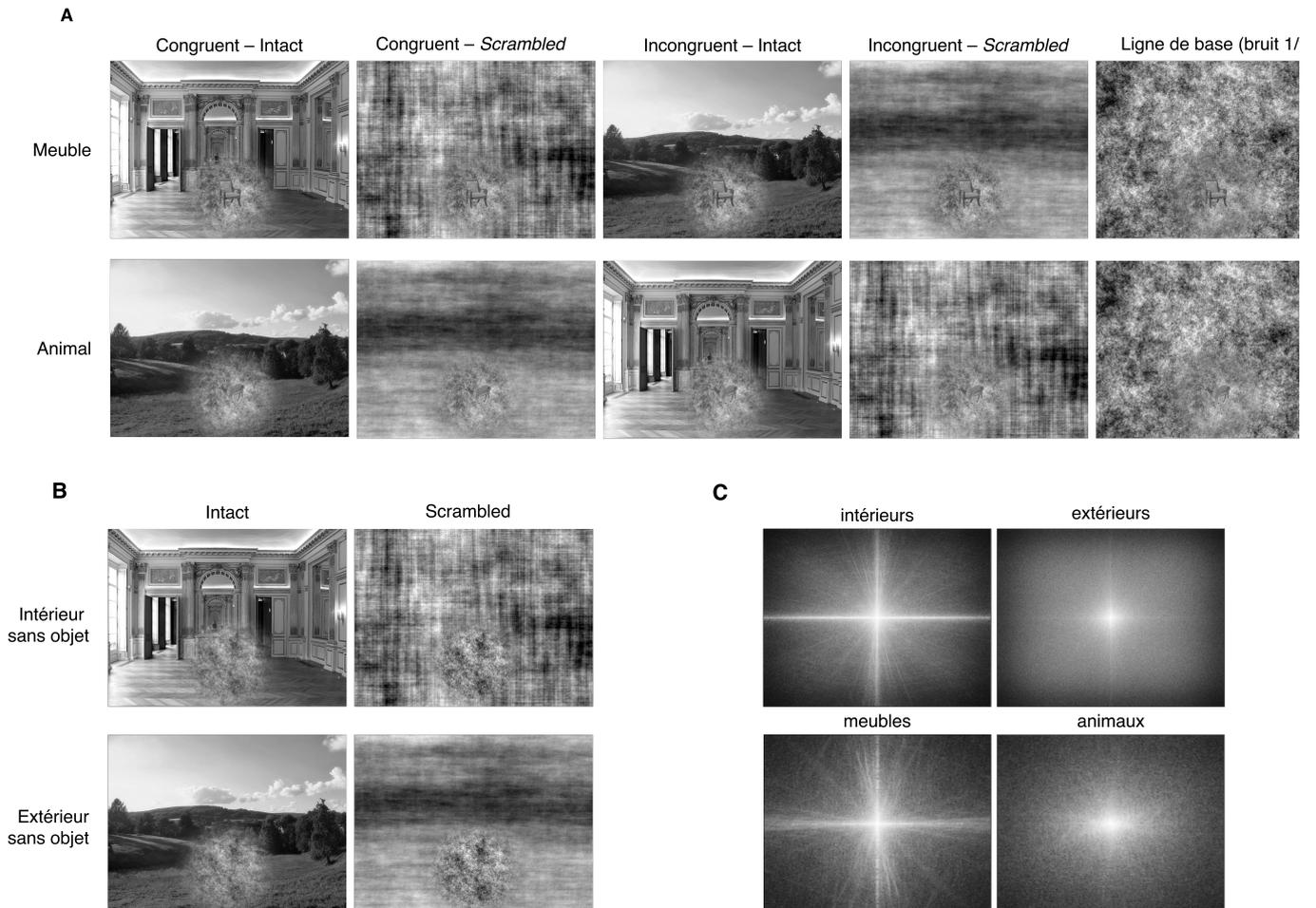


Figure 2.19: **A.** L'objet était un meuble (première ligne) ou un animal (deuxième ligne) collé sur un bruit distribué en $1/f$ et intégré soit dans une scène d'arrière-plan congruente ou incongruente, dont la phase était intacte ou *scrambled*, soit dans un arrière-plan sans signification (bruit $1/f$ - condition ligne de base). **B.** Condition 'scène-seule', dans laquelle il n'y avait pas d'objet dans le bruit $1/f$. **C.** Le spectre d'amplitude moyen des scènes d'intérieur est similaire à celui des meubles, alors que le spectre d'amplitude moyen des scènes d'extérieur est similaire à celle des animaux. Pour l'illustration, le contraste et la cohérence de phase des objets ont été légèrement augmentés dans cette figure.

4.6.2. Stimuli

Comme dans l'expérience 4, nous avons utilisé 20 groupes de quatre associations scène-objet. Les stimuli ont été construits sous MATLAB. La procédure était la suivante. Nous avons d'abord appliqué une correction gamma sur chaque objet et l'avons collé sur une image constituée d'un bruit distribué en $1/f$. Nous avons fait varier la visibilité de l'objet en manipulant un paramètre de mélange de phase. Pour chaque objet, sept versions de l'image bruitée avec l'objet étaient donc créés en diminuant la cohérence de la structure de phase de l'objet, en utilisant la fonction MATLAB de Ales, Farzin, Rossion, et Norcia (2012). Cette fonction interpole la phase entre la transformée de Fourier de l'image bruitée avec l'objet et une image de bruit blanc de même taille.

4. Expériences 4 et 5 : Influence de l'information contextuelle d'une scène en vision périphérique sur la catégorisation d'un objet en vision centrale

Les sept versions allaient de 0% à 80% de cohérence de phase (100% correspondant à l'image originale de l'objet), incrémentée progressivement par écarts d'environ 13% (résultant en sept versions de chaque objet de respectivement 0%, 13%, 27%, 40%, 54%, 67% et 80% de cohérence de phase). Nous avons choisi de limiter la cohérence maximale à 80% afin de minimiser les chances que les performances plafonnent comme dans l'Expérience 4 avec une pleine visibilité des objets. Chacune des sept versions des images 1/ f avec l'objet allait ensuite être intégrée sur cinq arrière-plans différents (intact-congruent, intact-incongruent, *scrambled*-congruent, *scrambled*-incongruent, bruit 1/ f). Les valeurs de luminance des scènes intactes étaient modifiées pour linéariser indirectement la fonction gamma de l'écran, puis égalisées pour obtenir une luminance moyenne de chaque scène de 0.51, et un RMS moyen de 0.24. Les images de scènes *scrambled* ont été créées en mélangeant la phase des scènes intactes dans le domaine de Fourier par permutation aléatoire. L'arrière-plan bruité de la condition ligne de base était un bruit 1/ f de la même taille que les scènes (i.e. 1024×768 pixels, ou $24 \times 18^\circ$). Les images d'objets étaient donc progressivement intégrées aux différents arrière-plans avec un patch circulaire (Figure 2.19). Pour la mesure des réponses précipitées, nous avons aussi créé des stimuli dans lesquels les scènes (intactes et *scrambled*) étaient présentées sans objet (l'endroit de l'image qui tombait en vision centrale était rempli de bruit 1/ f).

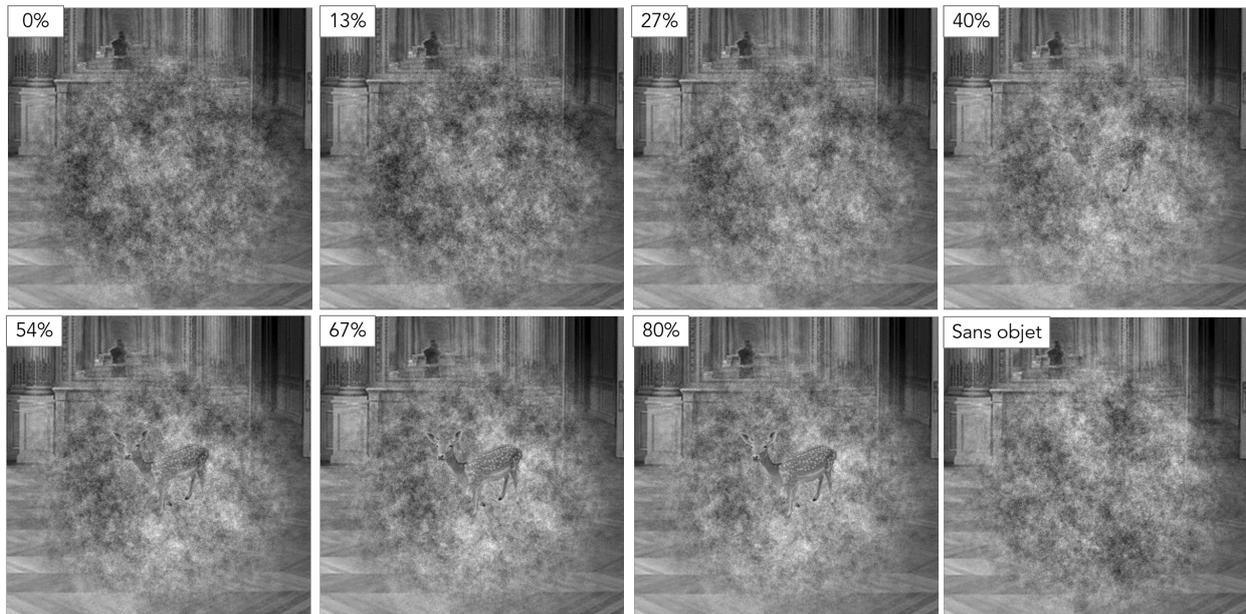


Figure 2.20: Portion d'une scène d'intérieur sur laquelle la vision centrale était dirigée, contenant un objet (un faon) ou pas d'objet. Les participants voyaient chaque objet dans sept versions allant de 0% à 80% de cohérence de phase. Dans un objectif d'illustration, le contraste et la cohérence de phase des objets ont été légèrement augmentés dans la figure par rapport à ceux des véritables stimuli.

4.6.3. Procédure

L'expérience était programmée sur E-Prime 2. Les stimuli étaient présentés sur le même écran 30 pouces que celui utilisé dans l'Expérience 4. À une distance de 50 cm, les scènes mesuraient 24° d'angle visuel horizontalement et 18° verticalement, le côté le plus grand des objets mesurait 3° , et le patch circulaire mesurait environ 8° . La tête des participants reposait sur une mentonnière pour maintenir la distance à l'écran. Ils avaient pour consigne de fixer une petite croix blanche centrée sur l'endroit de l'écran où l'objet allait être présenté (i.e. au centre de l'écran horizontalement et dans la partie basse verticalement). La croix de fixation était présentée pendant une durée qui variait aléatoirement entre 500 ms et 1500 ms (en moyenne environ 1000 ms) afin de perturber la prédictibilité du rythme des essais. Chaque essai commençait avec la petite croix de fixation, suivie par un premier stimulus (toujours un arrière-plan sans objet) pendant 150 ms, puis par un second stimulus pendant 150 ms, et enfin par un masque (nouveau bruit $1/f$) pendant 30 ms. Le second stimulus dépendait de la condition expérimentale : scène intacte avec un objet congruent, scène intacte avec un objet incongruent, scène *scrambled* avec un objet congruent, scène *scrambled* avec un objet incongruent, arrière-plan bruité avec un objet, scène intacte sans objet, et scène *scrambled* sans objet. Le premier stimulus était donc l'arrière-plan du second stimulus. L'intervalle interstimulus était de 3000 ms en moyenne. Puisque dans le second stimulus, il y avait toujours un patch circulaire bruité en vision centrale (qui comprenait un objet ou pas), pour maintenir un percept cohérent tout le long de l'essai, un patch circulaire bruité était aussi fondu dans le premier stimulus (Figure 2.21).

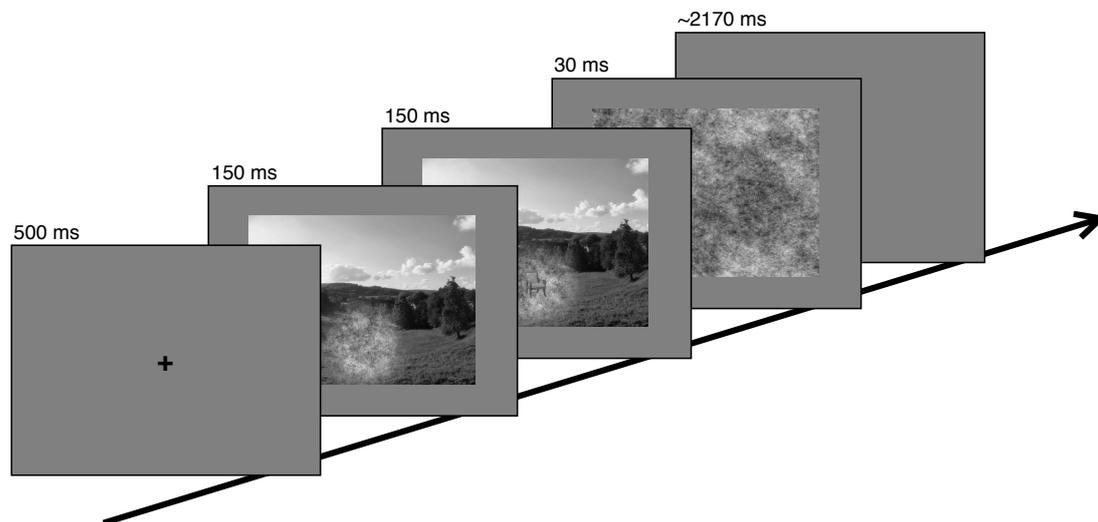


Figure 2.21: Procédure expérimentale de l'Expérience 5.

Comme dans l'Expérience 4, la tâche était de catégoriser l'objet selon qu'il s'agissait d'un animal ou d'un meuble en appuyant sur deux boutons d'un clavier. Puisque la visibilité de l'objet était basse dans beaucoup d'essais, nous encourageons les participants à se fier à leur intuition et ils avaient pour consigne de répondre sans favoriser l'un des

deux boutons de réponse (mais de toujours donner une réponse) lorsqu'ils ne pouvaient pas voir d'objet. Les associations entre les catégories et les boutons de réponse étaient contrebalancées entre les participants. L'exactitude de la réponse et le temps de réponse étaient enregistrés à chaque essai. Les conditions expérimentales étaient totalement randomisées. Il y avait 1720 essais : 280 essais scène intacte-congruent (20 animal/extérieur et 20 meuble/intérieur \times 7 niveaux de cohérence de phase de l'objet), 280 essais scène intacte-incongruent (20 animal/intérieur et 20 meuble/extérieur \times 7 niveaux de cohérence de phase), 280 essais scène *scrambled*-congruent (20 animal object/extérieur et 20 meuble/indoor \times 7 niveaux de cohérence de phase), 280 essais scène *scrambled*-incongruent (20 animal/intérieur et 20 meuble/extérieur \times 7 niveaux de cohérence de phase), 280 essais avec un arrière-plan bruité (ligne de base, 20 animal/fond bruité et 20 meuble/fond bruité \times 7 niveaux de cohérence de phase), 160 essais scène intacte seule (80 extérieures et 80 intérieures), 160 essais scène *scrambled* seule (80 extérieures et 80 intérieures). L'expérience était divisée en deux sessions d'une heure, comprenant les pauses. Avant de commencer chaque session, les participants réalisaient une session d'entraînement (20 essais) afin de se familiariser à la tâche et aux stimuli.

4.6.4. Traitement des données

Nous avons ajusté des fonctions Weibull liant la probabilité de réponses correctes de chaque participant au niveau de visibilité en estimant le maximum de vraisemblance pour les conditions congruente et incongruente des conditions intacte et *scrambled*, mais aussi pour la condition ligne de base, en utilisant le package *Quickpsy* dans R (Linares & López-Moliner 2016). Pour chaque condition expérimentale, nous avons ensuite extrait les seuils de visibilité des fonctions psychométriques, à savoir le niveau de cohérence de phase pour lequel la proportion de réponses correctes atteint 0.75 (1-p(hasard)/2). Les seuils de visibilité ont été analysés à l'aide de modèles linéaires à effets mixtes. À noter qu'il n'y a qu'une valeur de seuil par participant. Ces modèles mixtes prennent donc en compte l'effet aléatoire de l'intercept de chaque participant (i.e. son écart à la moyenne du groupe) mais pas l'effet aléatoire de la variabilité de chaque participant.

Les données utilisées pour l'analyse des temps de réponses correctes incluaient seulement les niveaux de visibilité (cohérence de phase) pour lesquels la proportion de réponses correctes moyenne excédait 0.75 au niveau du groupe, c'est-à-dire les niveaux 4, 5, 6 et 7, afin de moyenniser un nombre suffisant d'essais. Suite à la procédure d'écrémage, 0.35% des temps de réponse ont été retirés des données utilisées pour les analyses statistiques. Les seuils de visibilité moyens, les TRm et les logTRm, ainsi que les écart-types de chaque condition expérimentale, sont reportés dans le Tableau 2.6.

4.7. Expérience 5 : Résultats

Tableau 2.6: Résultats de l'Expérience 5. Moyennes \pm écart-types des seuils de visibilité (cohérence de phase), des temps de réponses correctes moyens en millisecondes (TRm), et des TR transformés (transformation logarithmique) moyens ($\log(\text{TR})_m$) en fonction de la Congruence sémantique (congruent ; incongruent) et de la Structure de phase de la scène (intacte ; *scrambled*).

	INTACT Congruent	INTACT Incongruent	SCRAMBLED Congruent	SCRAMBLED Incongruent	Niveau de base (bruit 1/f)
Seuil	3.44 \pm 0.65	3.76 \pm 0.70	3.59 \pm 0.52	3.62 \pm 0.68	3.57 \pm 0.51
TRm (ms)	522 \pm 57	550 \pm 54	541 \pm 62	533 \pm 58	533 \pm 52
$\log(\text{TR})_m$	2.70 \pm 0.06	2.73 \pm 0.04	2.72 \pm 0.05	2.71 \pm 0.04	2.71 \pm 0.04

4.7.1. Effet de congruence et nature des influences : scènes intactes vs scrambled

Les effets fixes des modèles mixtes ont été estimés pour la Congruence sémantique (congruent, incongruent), pour la Structure de phase de la scène périphérique (intacte, *scrambled*), et pour l'interaction de ces deux facteurs, et les effets aléatoires ont été pris en compte pour l'ordonnée à l'origine des participants. Les tests des effets simples ont été réalisés avec un ajustement du seuil alpha (correction de Bonferroni, $0.05/2 = 0.25$ pour les deux contrastes réalisés pour chaque variable dépendante).

L'analyse des seuils de visibilité montre un effet principal de la Congruence sémantique, avec des seuils plus bas lorsque la scène est congruente que lorsqu'elle est incongruente (congruent = 3.52 ± 0.58 , incongruent = 3.69 ± 0.68 ; $\beta = 0.17$, %change = 19.11%, $t(53) = 2.39$, $p = .021$; Figure 2.22-A et -C). Le modèle d'interaction montre que la structure de Phase de la scène interagit avec la Congruence ($\beta = -0.29$, $t(51) = -2.04$, $p = .047$; Fig. 5). Les modèles d'effets simples montrent un effet d'interférence lorsque l'information de phase des scènes est intacte (congruent = 3.44 ± 0.65 , incongruent = 3.76 ± 0.70 ; $\beta = 0.32$, %change = 37.89%, $t(17) = 2.74$, $p = .014$) alors qu'il n'y a pas de différence entre les conditions congruente et incongruente lorsque l'information de phase des scènes est *scrambled* ($\beta = 0.03$, $t(17) < 1$).

L'analyse des $\log\text{TR}_m$ ne révèle pas d'effet significatif de la Congruence sémantique ($\beta = 0.009$, $t(53) = 1.87$, $p = .062$). En revanche, le modèle d'interaction montre que la structure de Phase de la scène interagit avec la Congruence ($\beta = -0.04$, $t(51) = -3.16$, $p < .001$; Figure 2.22-B)). Les modèles d'effets simples montrent un effet d'interférence lorsque l'information de phase des scènes est intacte, les objets sont catégorisés plus rapidement lorsque la scène intacte est congruente que lorsqu'elle est incongruente (congruent = 522 ± 57 ms, incongruent = 550 ± 54 ms; $\beta = 0.03$, %change = 2.55%, $t(17) = 3.38$, $p < .001$), alors qu'il n'y a pas de différence entre les deux conditions lorsque l'information de phase des scènes est *scrambled* ($\beta = -0.007$, $t(17) < 1$).

4. Expériences 4 et 5 : Influence de l'information contextuelle d'une scène en vision périphérique sur la catégorisation d'un objet en vision centrale

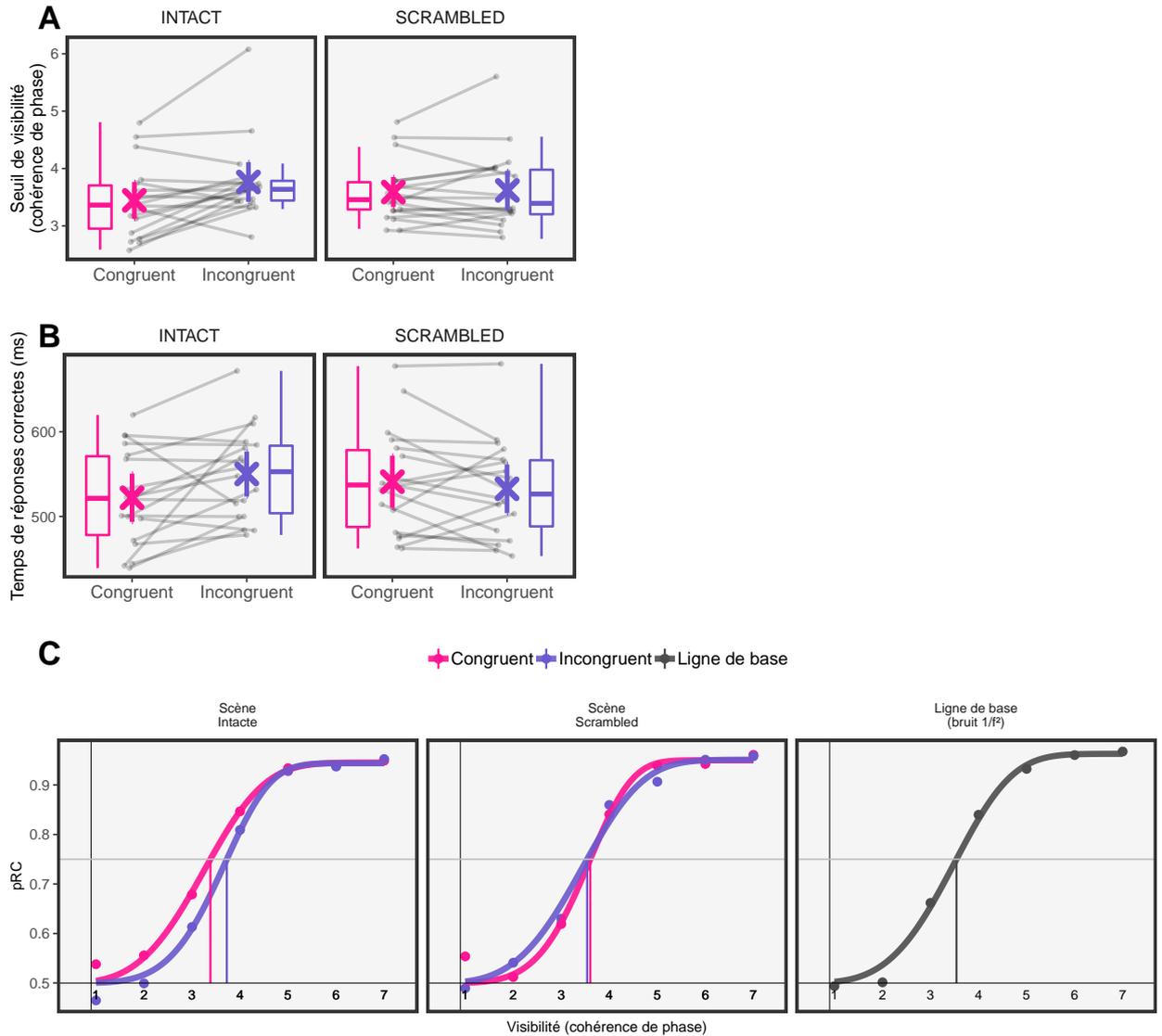


Figure 2.22: Résultats de l'Expérience 5. **A.** Seuils de visibilité (cohérence de phase) et **B.** Temps de réponses correctes (ms) en fonction de la Congruence sémantique (congruent; incongruent) et de la Structure de phase de la scène périphérique (intacte; *scrambled*). Les \times représentent les moyennes et les barres verticales qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **C.** Modèles Weibull ajustés à la proportion de réponses correctes pour les stimuli congruents et incongruents des conditions intacte et *scrambled*, et pour la condition ligne de base (pour laquelle l'arrière-plan était un bruit distribué en $1/f$).

Ces résultats supportent en partie ceux de l'Expérience 4 en suggérant que le niveau de visibilité nécessaire pour reconnaître un objet en vision centrale peut être diminué par la présence d'une information prédictive en vision périphérique. Par contre, contrairement à l'Expérience 4, nous n'avons pas observé l'effet de congruence sur les TRs. Par ailleurs, il semblerait qu'il faille que l'information visuelle périphérique soit porteuse

de sens pour influencer le traitement de l'objet, puisque nous n'observons pas d'effet de congruence lorsque la structure de phase de la scène périphérique est *scrambled*. Puisqu'il n'y a pas d'effet de congruence avec les scènes *scrambled*, les analyses suivantes considèrent seulement les conditions expérimentales pour lesquelles la structure de phase de la scène périphérique est intacte.

4.7.2. Direction de l'effet de congruence : Comparaison avec la ligne de base

Pour évaluer la direction de l'effet de congruence (facilitation de la catégorisation dans la condition congruente, gêne de la catégorisation dans la condition incongruente, ou les deux), nous avons comparé les seuils de visibilité et les logTRm de la condition niveau de base (fond bruité en $1/f$) à ceux des conditions congruente et incongruente en utilisant des modèles linéaires mixtes avec une correction de Bonferroni ($0.05/2 = 0.025$ pour les deux tests réalisés sur chaque variable dépendante). Les modèles estiment les effets fixes du type d'information en vision périphérique (scène congruente *vs* niveau de base; scène incongruente *vs* niveau de base), et les effets aléatoires sont pris en compte pour l'ordonnée à l'origine des participants.

Pour les seuils de visibilité, ces analyses ne montrent pas de différence significative, ni entre le niveau de base et la condition congruente ($\beta = 0.12$, $t(17) = 1.37$, $p = .190$), ni entre le niveau de base et la condition incongruente ($\beta = 0.20$, $t(17) = 1.99$, $p = .063$; Figure 2.23-A)). Pour les logTRm, il n'y a pas de différence entre le niveau de base et la condition congruente ($\beta = -0.009$, $t(1256) = -1.19$, $p = .233$) mais les participants étaient significativement plus lents dans la condition incongruente que dans la condition niveau de base (congruent = 522 ± 57 ms, ligne de base = 533 ± 52 ms; $\beta = -0.02$, $\%change = 1.64\%$, $t(1280) = -2.63$, $p = .009$; Figure 2.23-B)).

4.7.3. Relation entre l'effet de congruence et l'effet des réponses 'précipitées'

À chaque essai pour lequel il n'y avait pas d'objet dans l'image, nous avons étiqueté chaque réponse *animal* quand une scène d'extérieur était présentée, et chaque réponse *meuble* quand une scène d'intérieur était présentée, comme étant des 'réponses précipitées'. Si les participants n'utilisent pas de stratégie de réponse précipitée, ils sont supposés répondre au hasard (théoriquement 50% de réponses *animal* et 50% de réponses *meuble*) dans les essais sans objet. La proportion de réponses précipitées représente alors la tendance de chaque participant à utiliser la catégorie de la scène pour répondre. Nous avons d'abord testé si la proportion de réponses précipitées, ci-après nommée *index de réponses précipitées*, était différente de la chance (i.e. d'une proportion de 0.5) en utilisant un test de Student à échantillon unique. Même si la proportion moyenne de réponses précipitées était très proche du hasard (0.54 ± 0.04), elle était significativement différente du hasard ($t(17) = 4.28$, $p < 0.001$), indiquant que les réponses des participants étaient en partie liées à la catégorie de la scène (donc des réponses précipitées). Ensuite,

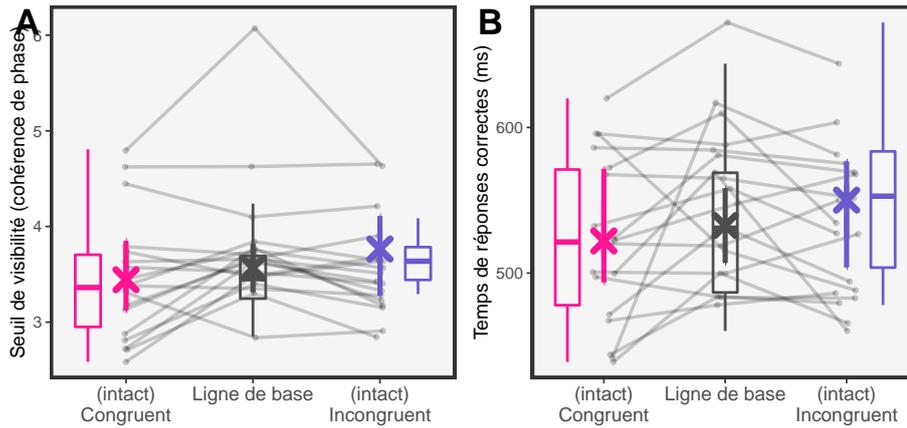


Figure 2.23: Résultats de l'Expérience 5. Comparaison avec la ligne de base (arrière-plan constitué d'un bruit distribué en $1/f$). **A.** Seuils de visibilité (cohérence de phase) et **B.** Temps de réponses correctes (ms) en fonction du type de scène en vision périphérique (congruente; incongruente; bruit $1/f$ -ligne de base). Les \times représentent les moyennes et les barres verticales qui s'y superposent sont les intervalles de confiance à 95%. Les droites correspondent à la pente de l'effet d'interférence pour chaque participant. **C.** Modèles Weibull ajustés à la proportion de réponses correctes pour les stimuli congruents et incongruents des conditions intacte et *scrambled*, et pour la condition ligne de base (pour laquelle l'arrière-plan était un bruit distribué en $1/f$).

nous avons testé une corrélation de Kendall entre l'index de réponses précipitées et l'effet d'interférence ([condition incongruente] - [condition congruente], dans la condition scène intacte). Il n'y avait pas de corrélation significative entre ces deux variables ($r = .16$, $p = .362$; Figure 2.24). Pour éviter de conclure sur l'hypothèse nulle, nous avons testé cette corrélation avec une approche bayésienne en utilisant le package *Psycho* dans R (Makowski 2018). Le modèle bayésien estime que la corrélation peut être considérée comme grande avec une probabilité de 2.94%, comme modérée avec une probabilité de 15.58%, comme petite avec une probabilité de 32.78% et comme très petite avec une probabilité de 18.25%. Le coefficient de corrélation peut alors être considéré comme petit, et l'analyse indique par ailleurs qu'il y aurait plutôt une absence d'association positive entre les deux variables ($r = 0.11$, écart absolu médian = 0.21, intervalle de crédibilité à 90 % [-0.23, 0.46]), mais avec un rapport de vraisemblance anecdotique ($BF = 1.76$).

4.8. Discussion des Expériences 4 et 5

Les Expériences 4 et 5 avaient pour objectif d'étudier comment une information de contexte disponible en vision périphérique pouvait influencer la reconnaissance d'objets en vision centrale en utilisant un paradigme d'interférence. Les participants catégorisaient des petits objets présentés en vision centrale tandis qu'une information de contexte congruente ou incongruente était perçue en vision périphérique.

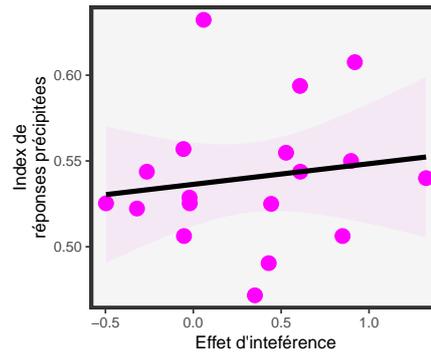


Figure 2.24: Résultats de l'Expérience 5. Relation entre l'effet d'interférence (différence de seuils de visibilité entre les conditions scène intacte congruente et scène intacte incongruente) et l'index de réponses précipitées (la tendance des participants à utiliser la catégorie de la scène pour inférer la catégorie de l'objet, lors des essais pour lesquels il n'y avait pas d'objet dans le stimulus). L'ombre représente l'intervalle de confiance à 95%.

Influence contextuelle périphérique et spécification du mécanisme prédictif

Tout d'abord, nous avons observé un effet d'interférence dans les deux expériences, indiquant que l'information de contexte en vision périphérique peut être traitée automatiquement et influencer les traitements réalisés en vision centrale. Par ailleurs, dans l'Expérience 5, l'analyse des seuils de visibilité a permis de montrer que la visibilité nécessaire pour catégoriser correctement un objet était plus basse lorsque le contexte périphérique était congruent que lorsqu'il était incongruent. Des mécanismes prédictifs utilisant l'analyse rapide de l'information visuelle périphérique pourraient alors booster le traitement perceptif de petits objets peu visibles dans la scène. De tels mécanismes peuvent certainement être un avantage pour la reconnaissance visuelle dans des conditions de faible éclairage ou de masquage. L'avantage de la mesure des seuils de visibilité est que les résultats permettent d'appuyer l'hypothèse que l'influence périphérique s'exerce sur un niveau perceptif plutôt que sémantique ou décisionnel. Autrement dit, l'influence périphérique (dans le cas d'une prédiction congruente) pourrait permettre de compléter le percept et de compenser l'information perceptive manquante. La reconnaissance de l'objet serait alors possible à un niveau de visibilité pour lequel elle serait impossible sans cette influence prédictive périphérique.

D'autre part, les résultats de l'Expérience 4 permettent de préciser les effets de contexte et de mieux les inscrire dans un cadre théorique de mécanismes prédictifs. Pour rappel, on peut considérer que les effets de contexte sur la reconnaissance d'objets peuvent être expliqués par deux hypothèses (Henderson & Hollingworth 1999). La

première suppose que l'objet et le contexte sont traités de manière principalement indépendante, en parallèle et la seconde, qui correspond au modèle FP-peri, suppose qu'il y a une précedence de traitement du contexte (représentation *gist*) par rapport à celui de l'objet, et que celui ci influencerait le traitement de l'objet *via* des mécanismes prédictifs. Dans l'Expérience 4, nous avons manipulé le temps de présentation de la scène avant l'onset de l'objet, ce qui nous a permis de contrôler la robustesse de la représentation *gist* du contexte. Nous avons observé que l'effet d'interférence sur les temps de réponse augmentait linéairement avec le SOA. Autrement dit, plus la représentation de la scène était stable et robuste, et moins il ne fallait de temps au participant pour catégoriser l'objet. Ce résultat est difficilement compatible avec l'hypothèse d'un traitement parallèle de l'objet et du contexte, puisque sous cette hypothèse, on aurait du observer le même effet d'interférence quel que soit le temps pendant lequel la scène pouvait être traitée avant l'onset de l'objet. Cette manipulation expérimentale simple permet ainsi d'appuyer l'hypothèse de mécanismes prédictifs. Dans le cadre du modèle FP-peri, l'information contextuelle portée par les BFs en vision périphérique pourrait ainsi alimenter rapidement le système de gestion des prédictions et influencer ainsi la reconnaissance de l'objet (cf. Figure 1.20 dans le Chapitre 1). À noter toutefois qu'avec une présentation simultanée de la scène et de l'objet (i.e. condition SOA-0ms), nous n'avons pas observé d'effet d'interférence. Ce résultat contraste avec l'effet d'interférence observé dans les Expériences 2 et 3, où les scènes périphérique et centrale étaient présentées simultanément. Cela dit, nous avons constaté un effet plafond dans l'Expérience 4. Dans cette expérience, l'entrée sensorielle est très bonne et la tâche est très facile (possiblement plus que la tâche de catégorisation de scènes dans les Expériences 2 et 3), ce qui a pu minimiser le poids des prédictions. Il est possible qu'un effet d'interférence soit mesurable en présentation simultanée de la scène et de l'objet lorsque, par exemple, les objets sont peu visibles - comme dans l'Expérience 5.

Stratégies de réponses dans le paradigme de présentation séquentielle

Comme souligné précédemment, l'utilisation d'un paradigme de présentation séquentielle peut soulever un problème méthodologique. Pour rappel, dans ce paradigme, l'effet de la congruence sur la catégorisation de l'objet peut être la conséquence de deux processus. Le premier serait que la catégorisation rapide et implicite de la scène entraîne automatiquement la réponse (stratégie de réponse précipitée). Le second serait une intégration de l'information visuelle périphérique aux processus de reconnaissance de l'objet en vision centrale. D'abord, nous avons observé qu'une partie des réponses des participants était en effet des réponses précipitées, puisque lorsque aucun objet n'était présent en vision centrale, les participants ne répondaient pas tout à fait au hasard ($0.54\% \pm 0.04$ des réponses étaient consistantes avec la catégorie de la scène, et cette proportion était significativement différente du hasard). Notons que ces réponses précipitées peuvent être liées à des processus décisionnels directement en lien avec le traitement de la scène (e.g., la reconnaissance d'une scène d'extérieur entraîne automatiquement et involontairement une catégorisation 'animal'), mais elles peuvent également refléter un phénomène

de percept illusoire dans le bruit, induit par le traitement de la scène. D'ailleurs, quasiment tous nos participants reportaient verbalement avoir des percepts illusoires d'objets n'appartenant pas à la base de stimuli (e.g., l'illusion de percevoir une tortue dans le bruit qui pourrait être induite par une scène de plage), et nous avons pu nous-mêmes faire l'expérience de ce phénomène. Notons que des effets similaires ont été reportés dans un contexte théorique de codage prédictif. Par exemple, [F. W. Smith et Muckli \(2010\)](#) ont mesuré une activation du cortex visuel rétinotopique qui n'était pas liée à l'input visuel (il n'y avait pas d'input visuel dans la partie correspondante du champ visuel) mais qui reflétait au contraire une information sur le contexte visuel. Néanmoins, quelle que soit l'origine des réponses précipitées des participants, nous n'avons pas obtenu de preuve solide d'un lien entre l'effet d'interférence et l'index de réponse précipitée, ce qui suggère que l'effet de la congruence sur la catégorisation des objets mesuré dans nos expériences est principalement dû à l'intégration de l'information visuelle périphérique aux processus de reconnaissance de l'objet en vision centrale (plutôt qu'à des réponses précipitées).

Sens de l'effet de congruence

Une question qui se pose avec le paradigme d'interférence concerne le sens de l'effet de la congruence sémantique. Dans l'Expérience 5, nous avons intégré une condition ligne de base pour laquelle l'information en vision périphérique était seulement du bruit (ni congruente ni incongruente). Les analyses montrent que les seuils de visibilité pour la condition ligne de base se situaient approximativement entre les seuils de la condition congruente et ceux de la condition incongruente (voir [Figure 2.23](#)) - même si la condition ligne de base n'était pas différente significativement ni de l'une, ni de l'autre. Pour les temps de réponse, nous avons cette fois-ci observé une différence entre la condition incongruente et la ligne de base. Globalement, le pattern de résultat n'est pas évident et, en tant que tel, suggère plutôt qu'un contexte incongruent retarde la catégorisation. Cette interprétation suggère que la décision de catégorisation serait retardée par un temps d'aller-retours entre le système de prédictions et le système de reconnaissance du modèle FP-péri, afin de résoudre l'erreur de prédiction et parvenir à un percept stable. À noter que nous n'avons donc pas répliqué l'effet de facilitation de la congruence observé par [Brandman et Peelen \(2017\)](#) (cf. Chapitre 1, sous-partie [3.2](#)). Cependant, la condition contrôle (ligne de base) dans leur étude consistait simplement à coller l'objet sur un fond gris uniforme. La comparaison de cette condition contrôle avec la condition ligne de base de notre Expérience 5 n'est donc que limitée, puisque nous avons intégré l'objet dans une image bruitée qui suit la relation entre amplitude et fréquence spatiale typique des scènes naturelles ($1/f$; [Ruderman & Bialek 1994](#)). Il est donc possible que la simple présence d'un bruit en vision périphérique ait boosté la perception de l'objet dans la condition ligne de base, par exemple *via* des mécanismes de résonance stochastique ([McDonnell & Ward 2011](#)).

Nature de l'influence contextuelle périphérique

Un autre objectif de l'Expérience 5 était de mieux caractériser la nature de l'influence périphérique. Nous avons manipulé la possibilité d'extraire une représentation sémantique de la scène périphérique, en présentant chaque scène dans sa version intacte (propriétés de bas niveau + représentation sémantique) et dans sa version *scrambled* (propriétés de bas niveau seulement). Les analyses ne montrent aucun effet d'interférence lorsque la scène est *scrambled* et suggèrent alors qu'il ne peut pas y avoir d'effet d'interférence lorsque seules les propriétés de bas niveau des scènes sont disponibles. La possibilité d'extraire une représentation de plus haut niveau semble nécessaire pour que l'information contextuelle en vision périphérique puisse guider les mécanismes de reconnaissance de l'objet en vision centrale. Pourtant, une limite à l'approche méthodologique que nous avons adoptée est que l'on ne peut pas affirmer que l'information qui provoque l'effet d'interférence avec les scènes intactes est une information proprement sémantique/abstraite. Il se peut que l'information importante soit une information sur la *localisation/configuration spatiale* de l'information visuelle. Selon cette hypothèse, ce n'est pas l'impossibilité d'extraire une représentation sémantique qui empêcherait l'effet d'interférence dans la condition scènes *scrambled*, mais l'impossibilité d'avoir accès à la configuration spatiale des scènes à cause de la perturbation de la phase de l'image. Ainsi, il est possible que le mélange de phase ne soit pas un si bon modèle des processus visuels de bas niveau, et que la distribution spatiale des caractéristiques de bas niveau soit en fait importante.

En utilisant le modèle computationnel HMAX (Riesenhuber & Poggio 2002, voir aussi <https://maxlab.neuro.georgetown.edu/hmax.html>), conçu pour simuler l'activité neurale à différentes étapes du traitement visuel et basé sur les propriétés physiologiques et fonctionnelles du cortex visuel, Stojanoski et Cusack (2014) ont testé la validité de la méthode du mélange de phase (*phase scrambling*) comme modèle des processus visuels de bas niveau. Plus précisément, ils ont prédit l'activité neurale des régions visuelles précoces induite par des images intactes, par des images dont la phase était *scrambled*, et par des images diffeomorphes. Les images diffeomorphes sont des images sans signification semblant être imprimées sur un tissu distordu (Figure 2.25). Cette étude montre que selon le modèle HMAX, ces régions visuelles précoces ne répondraient pas aux stimuli *phase scrambled* de la même manière qu'ils répondraient aux stimuli intacts. Par contre, l'activité neurale de ces régions induite par des stimuli diffeomorphes ne pourrait pas être distinguée de l'activité neurale induite par des images intactes. Les stimuli diffeomorphes pourraient ainsi être une meilleure manière de modéliser les processus visuels de bas niveau.

Ainsi, les résultats de l'Expérience 5 permettent d'avancer que des scènes périphériques *scrambled*, contenant seulement la distribution de l'amplitude sur les différentes orientations et fréquences spatiales, ne permettent pas de former des prédictions qui influencent la reconnaissance d'objet en vision centrale. Cependant, il est possible que d'autres propriétés visuelles de bas niveau, comme celles contenues dans des images diffeomorphes, puissent apporter une information suffisante pour former des prédictions.

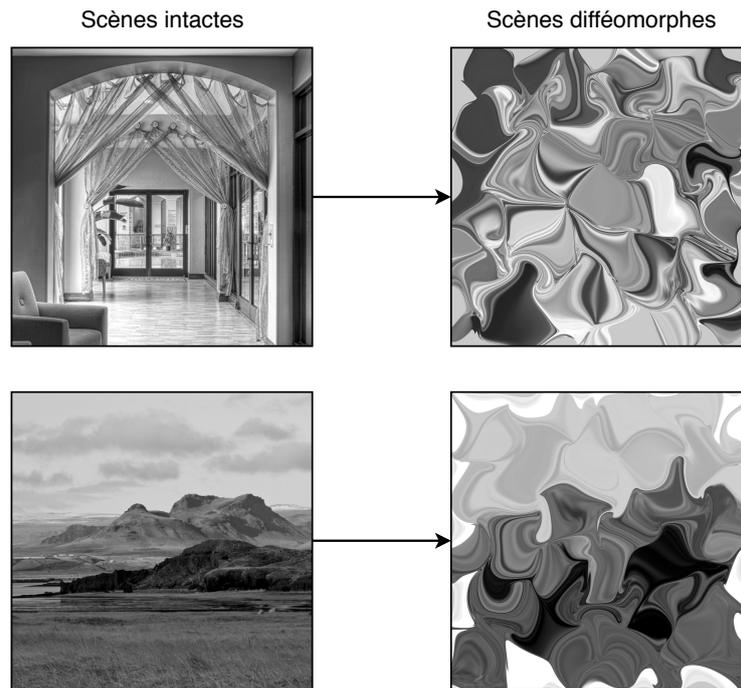


Figure 2.25: Scènes intactes et leurs versions diffeomorphes, d'après l'algorithme de [Stojanoski et Cusack \(2014\)](#) .

On ne peut donc pas conclure que l'accès à une représentation sémantique/abstraite est requise pour former des prédictions influençant les traitements en vision centrale. Une perspective pourrait être de comparer l'effet d'interférence induit par des scènes intactes, des scènes *scrambled* et des scènes diffeomorphes. En parallèle, on peut aussi souligner qu'il est difficile de conclure sur l'hypothèse nulle (absence d'effet dans la condition *scrambled*). Il est possible que l'information contenue dans les scènes *scrambled* puisse permettre d'influencer le traitement de l'objet dans une moindre mesure, mais que nous n'ayons pas pu mesurer cet effet ici à cause d'un manque de puissance statistique. En effet, nous avons planifié notre taille d'échantillon pour l'Expérience 5 en nous basant sur la taille de l'effet d'interférence de l'Expérience 4, où seules des scènes intactes étaient présentées.

Chapitre 3

Contribution empirique :

*Effet d'une dégradation de la vision
périphérique sur les traitements réalisés
en vision centrale chez le patient atteint
de glaucome*

1. Contexte et hypothèses

Le glaucome est la deuxième cause de malvoyance dans le monde, avec plus de 50 millions de personnes concernées (Quigley & Broman 2006), la première étant la cataracte, qui est réversible. Le glaucome primitif à angle ouvert (GPAO) est la forme la plus courante de glaucome. Le GPAO est caractérisé par une gêne de l'écoulement de l'humeur aqueuse de la chambre antérieure de l'oeil, du fait d'une altération progressive du filtre d'évacuation (le trabéculum). L'augmentation du liquide à l'intérieur de l'oeil est à l'origine d'une pression intraoculaire élevée liée à la destruction progressive du nerf optique et des cellules ganglionnaires rétiniennes (Fechtner & Weinreb 1994). Des analyses histologiques réalisées auprès de patients GPAO montrent que les neurones avec des corps cellulaires et des diamètres axonaux plus grands sont plus à risque d'être affectés au début du processus pathologique (Glovinsky, Quigley, & Dunkelberger 1991; Quigley, Sanchez, Dunkelberger, L'Hernault, & Baginski 1987), suggérant une vulnérabilité spécifique des cellules rétiniennes magnocellulaires. Les cellules parvocellulaires seraient également atteintes à des stades plus avancés de la maladie (Glovinsky et al. 1991; Quigley et al. 1987). Le GPAO cause une perte progressive mais inéluctable de la vision qui se caractérise le plus souvent par une perte progressive du champ visuel périphérique progressant peu à peu vers le centre.

1.1. Atteinte de la vision périphérique dans le glaucome

Dans la pratique clinique, la perte de champ visuel (CV) est évaluée à l'aide d'une périmétrie statique automatisée, un examen basé sur la détection de points lumineux. Cet examen teste le seuil de sensibilité rétinienne au contraste de luminance sur différents points dans les 10, 24 ou 32 degrés centraux du CV (selon le protocole utilisé). Le test est réalisé avec la correction optique habituelle du patient, de façon monoculaire ou binoculaire. L'évaluation est basée sur l'incrément progressive de la luminance de petits points lumineux projetés sur une coupole, à différentes excentricités rétiniennes. Le patient doit garder le regard fixé sur une cible au centre de la coupole et appuyer sur un bouton avec sa main dès qu'il détecte un point lumineux (Figure 3.1). On évalue ainsi le seuil de détection, exprimé en décibels (dB; i.e. transformation logarithmique de valeurs brutes de luminance en apostilbs), dans différentes régions du CV. L'examen fournit des informations diverses. La sensibilité de chaque point testé peut être évaluée (i) en calculant sa déviation par rapport à la sensibilité d'une population de référence (même tranche d'âge) sur le même point, c'est la déviation totale, ou (ii) en calculant sa déviation par rapport à la sensibilité de l'ensemble des points testés du patient, c'est la déviation individuelle. Ces informations peuvent aussi être résumées par des indices globaux. La déviation moyenne (*Mean Deviation*, MD) représente la moyenne des déviations de sensibilité du patient sur chaque point testé par rapport à la population de référence, et la déviation individuelle (*Pattern Standard Deviation*, PSD) représente le

degré de variation de l'ensemble du CV par rapport à la moyenne de tous les points testés du patient. Bien que l'indice MD ne permette pas de caractériser la position spatiale des scotomes dans le CV (un même indice MD peut être associé à un déficit localisé sévère, ou bien à un déficit moins sévère mais plus diffus), cet indice est souvent utilisé car il permet d'estimer rapidement la gravité de l'atteinte périmétrique.



Figure 3.1: Réalisation d'une périmétrie statique automatisée. Chaque oeil est testé séparément, avec si nécessaire une correction de la réfraction adaptée à la vision de près. Le patient est placé à 33 cm de la coupole d'examen, sur laquelle sont projetés les points lumineux. Le patient appuie sur un bouton de réponse à l'aide d'une manette lorsqu'il perçoit un point lumineux.

Une atteinte périmétrique est diagnostiquée dès lors qu'il y a une sensibilité anormale sur un des points testés. On parle au contraire de glaucome pré-périmétrique lorsqu'il n'y a pas encore de déficit décelé par l'examen périmétrique. Le diagnostic du glaucome peut dans ce cas être posé sur la base d'autres indices cliniques. L'importance de l'atteinte périmétrique peut être déterminée grâce à la classification HPA (Hodapp, Parrish, & Anderson 1993). Cette classification est notamment basée sur l'indice MD, bien qu'elle considère aussi le nombre de points testés présentant une sensibilité anormale, ou encore la présence d'une sensibilité anormale sur les quatre points centraux (voir Tableau 3.1).

Il suffit de jeter un oeil à l'examen du CV de quelques patients atteints de glaucome pour constater que la perte de CV est bien différente de celle d'une vision tubulaire dont les patients sont supposés faire l'expérience. De plus, des lésions précoces au centre de la rétine (i.e. au niveau de la macula) peuvent également être observées (Hood, Raza, de Moraes, Liebmann, & Ritch 2013), et la localisation des scotomes dans le champ visuel est très variable (voir Figure 3.2 pour des exemples). À cause de la répartition de

Tableau 3.1: Classification HPA (Hodapp et al. 1993) de la sévérité de l'atteinte périmétrique glaucomateuse.

Déficit mineur	Déficit modéré	Déficit sévère
MD < -6 dB	MD < -12 dB	MD > -12 dB
Moins de 18 points avec $p < 5\%$ et moins de 10 points testés avec $p < 1\%$	Moins de 37 points avec $p < 5\%$ et moins de 20 points testés avec $p < 1\%$	Plus de 37 points avec $p < 5\%$ et/ ou plus de 20 points testés avec $p < 1\%$
Absence de déficit supérieur à 15 dB dans les 5 degrés centraux	Absence de déficit absolu (0 dB) dans les 5 degrés centraux	Présence de déficit absolu (0 dB) dans les 5 degrés centraux
	Pas plus d'un hémichamp ayant des points déficitaires supérieurs à 15 dB dans les 5 degrés centraux	Présence de points dans les 5 degrés centraux avec des déficits supérieurs à 15 dB dans les 2 hémichamps.

leurs axones dans le nerf optique, les cellules ganglionnaires de la rétine qui présentent une sensibilité déficitaire ne sont pas réparties aléatoirement. Au contraire, le faisceau de fibres nerveuses endommagé est constitué des axones de cellules disposées de manière adjacente sur la rétine. Les scotomes dépendent ainsi de la configuration des fibres nerveuses abîmées. Au niveau phénoménologique, la perte de vision correspondrait davantage à des informations manquantes et à des distorsions floues plutôt qu'à une vision 'en tunnel' ou à des 'trous noirs' dans le CV (Crabb 2016; Crabb, Smith, Glen, Burton, & Garway-Heath 2013).

Comme le GPAO est caractérisé par la destruction des cellules rétinienne, les études sur le fonctionnement visuel des patients se focalisent sur des aspects de bas niveau du traitement visuel, comme la sensibilité au contraste évaluée avec la périmétrie statique automatisée ou avec d'autres tests (e.g., Pelli-Robson, Hawkins, Szlyk, Ardickas, Alexander, & Wilensky 2003; Hot, Dul, & Swanson 2008). Cependant, la perte de CV a des incidences sur des aspects du traitement visuel plus complexes que la simple détection de points lumineux.

1.2. Troubles visuels cognitifs en vision résiduelle

Typiquement, les patients rapportent des difficultés dans leur vie quotidienne même lorsque que leur atteinte périmétrique est seulement légère. Des travaux expérimentaux mettent en évidence les difficultés des patients pour atteindre et saisir des objets (Kotecha, O'Leary, Melmoth, Grant, & Crabb 2009), pour rechercher des objets (N. D. Smith, Crabb, & Garway-Heath 2011; N. D. Smith, Glen, & Crabb 2012), pour se déplacer (Ramulu 2009), ou encore pour exécuter des actions du quotidien (par exemple, préparer un sandwich Dive et al. 2016). L'atteinte de la sensibilité rétinienne, même légère, impacte ainsi des fonctions de plus haut niveau du traitement visuel, et peut se répercuter ainsi sur le fonctionnement global de l'individu. Ces dernières années, des chercheurs

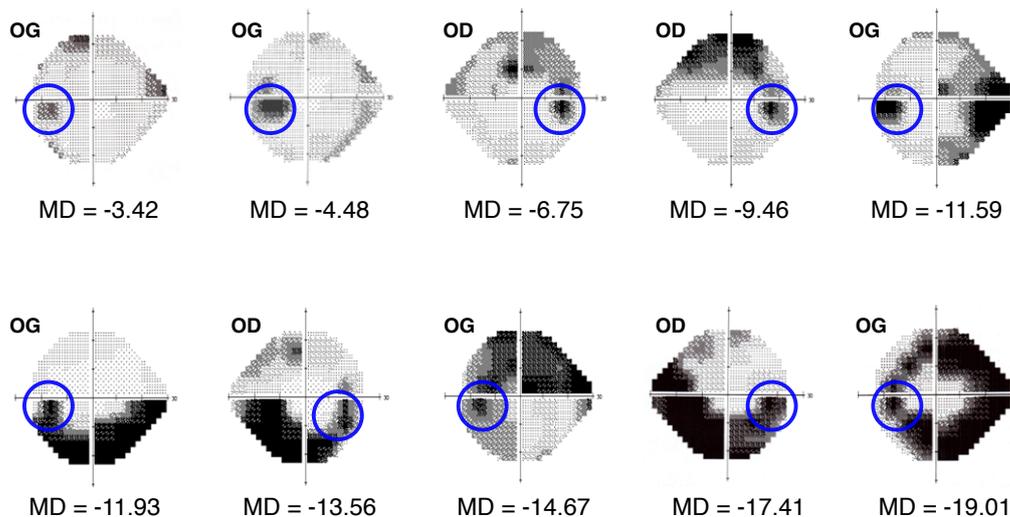


Figure 3.2: Exemple de ‘cartes’ de champs visuels monoculaires (OG : oeil gauche ; OD : oeil droit) de patients atteints de glaucome. Ces représentations sont issues d’examens de la périmétrie statique automatisée Humphrey SITA-Standard 24-2. Cet examen teste le seuil de sensibilité au contraste de 54 points du champ visuel, séparés de 6 degrés d’angle visuel dans les 24 degrés centraux des côtés supérieur, temporal et inférieur, et dans les 30 degrés centraux du côté nasal. Les points noirs indiquent une faible sensibilité au contraste. Comme le test est ici monocular, on distingue le point aveugle du côté temporal (entouré en bleu). Chaque représentation du champ visuel est associée à une valeur de *Mean Deviation* (MD) représentant la moyenne des déviations de sensibilité du patient sur chaque point testé par rapport à la population de référence.

ont commencé à évaluer les capacités des patients à réaliser des tâches visuelles de plus haut niveau que de la simple détection de contraste, et leurs résultats semblent pointer des déficits complexes, même dans leur vision résiduelle, c’est-à-dire dans les régions du CV classifiées comme normales par la périmétrie statique automatisée. Par exemple, sur la base de l’hypothèse d’une atteinte sélective de la voie magnocellulaire aux stades précoces de la maladie (Glovinsky et al. 1991; Quigley et al. 1987), McKendrick, Badcock, et Morgan (2004) ont étudié les mécanismes d’adaptation des voies magnocellulaire et parvocellulaire chez des patients GPAO. Les auteurs ont utilisé un paradigme de *steady/pulsed pedestal* adapté de Pokorny et Smith (1997) qui permet de tester sélectivement l’une des deux voies. Dans une première condition, que l’on appelle "piédestal continu" (*steady pedestal*, Figure 3.3-A), les auteurs présentaient d’abord aux patients quatre carrés (le piédestal) sur un fond gris fortement contrasté. Une fois le patient adapté à ce piédestal, un seul des carrés changeait de luminance. Le patient devait indiquer quel était le carré dont la luminance changeait. Il devait donc traiter des faibles différences de luminance entre le carré cible et le piédestal. Cette condition est supposée stimuler la voie magnocellulaire. Dans une seconde condition, qu’on appelle "piédestal flashé" (*pulsed pedestal*, Figure 3.3-B), les auteurs présentaient d’abord un fond gris. Une fois le patient adapté au fond gris, quatre carrés (ou piédestal) apparaissent dont

un de luminance différente. À nouveau, le patient devait indiquer quel était le carré dont la luminance changeait. Il pouvait dans cette condition comparer la luminance du carré cible à la luminance du fond de l'écran qui présentait un fort contraste. Cette condition est supposée stimuler la voie parvocellulaire. Pour chaque patient, le piédestal de chaque condition expérimentale était présenté dans un quadrant du CV périphérique dont la sensibilité rétinienne était évaluée comme normale par la périmétrie. Pour chaque condition, les auteurs mesuraient le seuil de détection du carré cible, c'est-à-dire la plus petite différence de luminance entre le carré cible et le piédestal pour laquelle le patient arrivait à détecter le carré cible. Ils ont ainsi montré que pour les deux conditions expérimentales, les seuils de détection des patients GPAO étaient plus élevés que ceux de participants du même âge avec une vision normale. Cette étude suggère qu'il y aurait un dysfonctionnement aussi bien de la voie parvocellulaire que de la voie magnocellulaire pour traiter le contraste dans la vision périphérique résiduelle des patients GPAO. Ce résultat irait à l'encontre d'une atteinte sélective de la voie magnocellulaire, tel que suggéré par les analyses histologiques menées auprès de patients glaucomateux (Glovinsky et al. 1991; Quigley et al. 1987). On peut toutefois remarquer que les auteurs ont testé des patients à des stades variés de la maladie. Certains patients présentaient un glaucome précoce tandis que d'autres présentaient un glaucome à un stade plus avancé. Il est donc possible que les résultats moyennés de cette étude masquent une atteinte sélective de la fonction magnocellulaire qui serait caractéristique des stades précoces de la maladie. Mais cette étude a surtout permis de mettre en évidence pour la première fois qu'il pouvait y avoir des déficits visuels subtils en vision résiduelle (i.e. dans des régions du CV considérées comme normales par l'examen de la périmétrie) lorsque le stimulus et/ou la tâche augmentent en complexité.

Par ailleurs, des fonctions visuelles dites cognitives seraient également altérées. Par exemple, McKendrick, Badcock, et Morgan (2005) ont testé des patients atteints de GPAO sur une tâche de *global motion*, dans laquelle l'objectif était de déterminer la direction globale d'un ensemble de points en mouvement, et sur une tâche de *global form* dans laquelle l'objectif était de déterminer si un ensemble de points ("patterns de Glass") formait un pattern radial ou un pattern concentrique (Figure 3.4-A). Les patients montraient de grosses difficultés à extraire ces patterns globaux. Des déficits complexes directement liés à la reconnaissance visuelle ont été observés pour des tâches de reconnaissance de visages et d'objets (Glen, Crabb, Smith, Burton, & Garway-Heath 2012; Lenoble et al. 2016; Schafer, Rouland, Peyrin, Szaffarczyk, & Boucart 2018). Les patients présentent également des mouvements oculaires atypiques au cours de la recherche visuelle d'objets (N. D. Smith et al. 2012) ou de l'exploration de visages (Glen, Smith, & Crabb 2013). Le plus surprenant est que certains de ces travaux montrent que de tels déficit existent même dans les régions visuelles résiduelles (Lenoble et al. 2016; McKendrick et al. 2005; Schafer et al. 2018). Par exemple, Lenoble et al. (2016) ont mis en évidence que des patients atteints de GPAO pour lesquels l'examen périmétrique n'indiquait aucune perte de sensibilité dans le CV central avaient de moins bonnes performance que des sujets contrôles pour catégoriser de petits objets flashés (28 ms) en vision fovéale

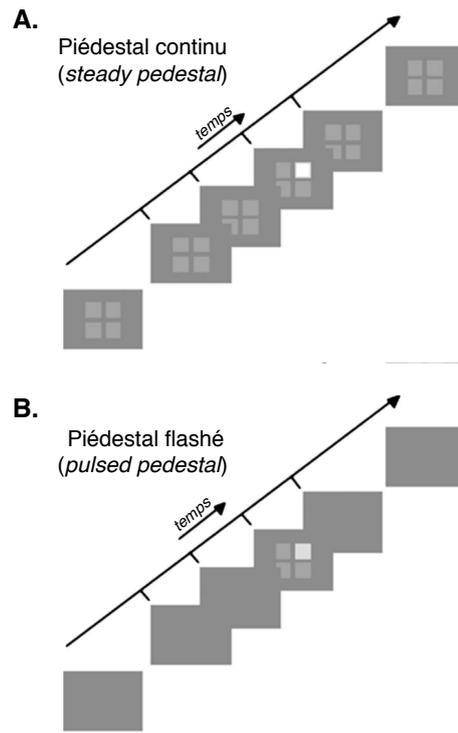


Figure 3.3: Illustration du paradigme des *Steady-Pulsed pedestal* utilisé par McKendrick et al. (2004). **A.** Condition piédestal continu. **B.** Condition piédestal flashé. Adapté de Pokorny et Smith (1997).

dès lors que le contraste était atténué de moitié. Les capacités de reconnaissance visuelle des patients en vision centrale pourraient donc être plus impactées que ce qui pourrait être prédit sur la base de l'évaluation de la perte de sensibilité rétinienne. Autrement dit, les processus visuels de moyen et haut niveau menant à la reconnaissance visuelle pourraient être altérés et/ou plus coûteux chez ces patients, même dans des régions du CV pour lesquelles la sensibilité rétinienne est normale. Malheureusement, encore trop peu de travaux ont été consacrés à l'étude de la cognition visuelle chez ces patients et à ce jour, aucun modèle théorique n'a été proposé pour rendre compte de leur déficits cognitifs.

1.3. Objectifs et hypothèses théoriques

Dans ce travail de thèse, nous faisons l'hypothèse d'un modèle FP-peri, selon lequel l'émergence rapide d'une première représentation *gist*, sur la base de l'extraction des BFs disponibles en vision périphérique, influencerait la reconnaissance visuelle en vision centrale à travers des mécanismes de prédictions. Suivant cette hypothèse, le système prédictif de patients atteints d'un glaucome serait alors perturbé. Plus précisément, le *gist* serait de moins bonne qualité chez ces patients à cause de l'altération irréversible de la rétine périphérique et donc de l'impossibilité à traiter les inputs qui s'y trouvent. Le système prédictif serait alors moins efficace, perturbant les processus de reconnaissance

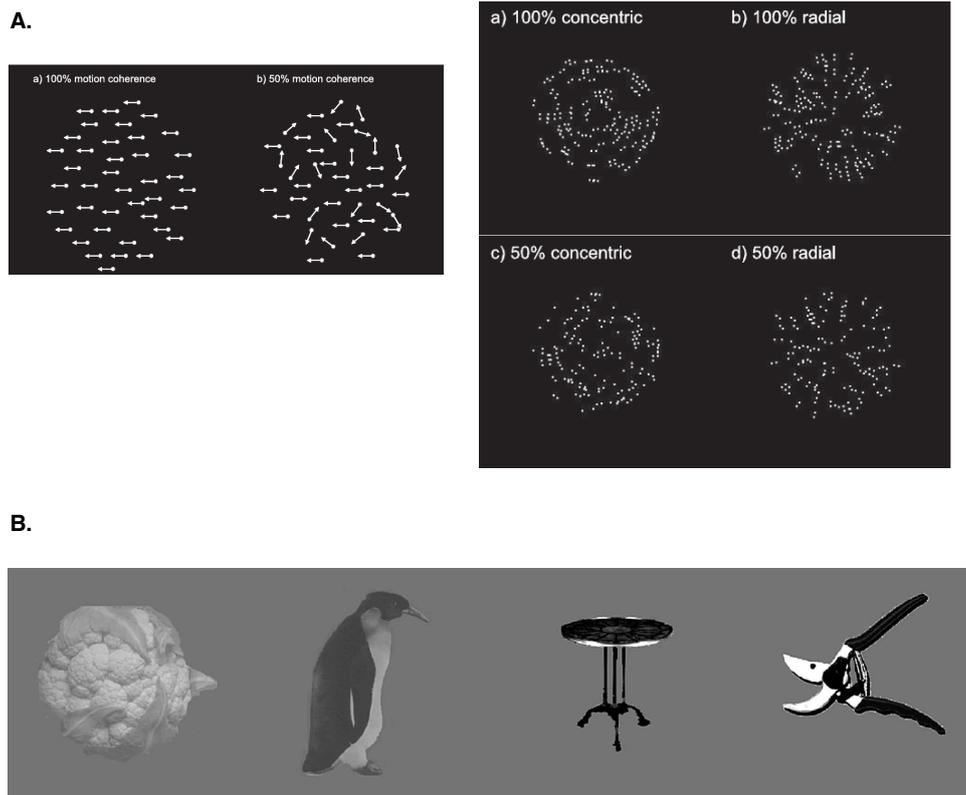


Figure 3.4: Stimuli utilisés pour l'étude des processus visuels cognitifs de patients atteints de glaucome. **A.** McKendrick et al. (2005) : Les patients réalisaient des tâches de discrimination de *global motion* (stimuli de gauche) et de *global form* (stimuli de droite). Les stimuli étaient présentés dans des quadrants du champ visuel considérés comme intacts par la périmétrie statique automatisée. **B.** Lenoble et al. (2016) : Les patients catégorisaient de petits objets présentés à 50% de contraste (stimuli de gauche) ou 100% de contraste (stimuli de droite). Les stimuli étaient présentés dans la vision centrale intacte des patients.

en vision résiduelle, et notamment en vision centrale. Une hypothèse assez similaire a déjà été avancée par Lenoble et al. (2016). Selon les auteurs, le glaucome affecterait la possibilité d'extraire une première représentation grossière et impacterait ainsi la reconnaissance des objets. Ces hypothèses sont cohérentes avec les résultats qui montrent que des patients présentent des déficits de reconnaissance visuelle en vision centrale alors même que cette partie du CV est diagnostiquée comme préservée au cours de l'examen périmétrique (Lenoble et al. 2016; McKendrick et al. 2005; Schafer et al. 2018). À cause de la perte progressive de stimulation de la rétine périphérique, ces patients ne bénéficieraient pas pleinement des processus prédictifs normalement utilisés pour traiter l'information visuelle en vision centrale. Nous pensons donc que le glaucome représente un bon modèle pathologique pour étudier le rôle de la vision périphérique dans la reconnaissance visuelle.

De plus, nous faisons l'hypothèse qu'un dysfonctionnement des mécanismes prédictifs inhérents à la vision périphérique impacterait tout particulièrement la reconnaissance

visuelle en vision centrale en particulier lorsque la stimulation visuelle est de faible qualité perceptive. En effet, comme nous en avons déjà fait état, l'intérêt d'un codage prédictif de l'information visuelle serait notamment de pouvoir mieux reconnaître les stimuli de faible qualité perceptive. Les résultats de l'Expérience 5 de ce travail de thèse suggèrent par exemple que la reconnaissance d'un stimulus de faible qualité perceptive est améliorée par la présence d'une information utile en vision périphérique. Si une altération irréversible de la vision périphérique par le glaucome gêne les processus prédictifs de reconnaissance visuelle en vision centrale, cette gêne devrait être d'autant plus forte que le stimulus visuel est partiellement masqué, brouillé, ou faiblement contrasté. Nous avons donc mené une expérience comportementale (**Expérience 6**) afin de tester les capacités de patients GPAO à catégoriser des stimuli visuels faiblement contrastés en vision centrale. Nous avons ensuite mené une étude exploratoire en IRMf (**Expérience 7**) afin d'étudier si une désorganisation du système prédictif des patients pourrait se manifester au niveau cérébral. Nous avons donc adapté le protocole de l'Expérience 6 à l'IRMf afin d'étudier de manière exploratoire les changements fonctionnels cérébraux chez trois patients glaucomateux et quatre participants contrôles appariés en âge lors de la catégorisation de scènes visuelles faiblement contrastées en vision centrale.

2. Expérience 6 : Étude de la reconnaissance de scènes et de visages faiblement contrastés dans la vision centrale de patients atteints d'un glaucome

N.B. Cette étude a été publiée dans la revue PLOS One ([Roux-Sibilon et al. 2018](#); Annexe C).

2.1. Objectifs et hypothèses opérationnelles

Dans l'Expérience 6, nous avons pour objectif d'étudier l'impact d'une altération de la vision périphérique sur la reconnaissance de stimuli visuels de faible qualité perceptive. La qualité perceptive d'un stimulus visuel peut être manipulée de diverses manières : brouillage d'une partie du spectre de phase (cf. Expérience 5), masquage partiel ou baisse du contraste de luminance. Ici, nous avons choisi de manipuler le contraste de luminance des stimuli. Ce choix méthodologique nous semble particulièrement approprié dans le cas du glaucome. En effet, l'une des principales plaintes des patients glaucomateux est liée aux conditions de luminosité. Les difficultés liées à l'éclairage telles que l'éblouissement et la difficulté à s'adapter à différents niveaux d'éclairage sont régulièrement classées parmi les plaintes les plus fréquentes ([Crabb et al. 2013](#); [B. L. Lee et al. 1998](#); [Nelson, Aspinall, & O'Brien 1999](#)). Par exemple, les patients expriment significativement plus de difficultés à voir dans l'obscurité (i.e. faible luminosité) et à la lumière du jour (i.e. forte luminosité) que les personnes non atteintes ([B. L. Lee et al. 1998](#)). Dans ces deux situations, le contraste de luminance est faible. Les patients rapportent également des phénomènes d'éblouissement, de halos lumineux autour des sources lumineuses et des difficultés à s'adapter lors de changements de luminosité ([Dubey et al. 2019](#); [B. L. Lee et al. 1998](#); [Nelson et al. 1999](#)). Une mauvaise sensibilité au contraste chez les patients atteints de glaucome est plutôt bien vérifiée expérimentalement, en utilisant des tests comme les optotypes de Pelli-Robson, ou des paradigmes plus complexes comme les *steady/pulsed pedestals* de [Pokorny et Smith \(1997\)](#) déjà évoqués plus haut (voir par exemple [Hawkins et al. 2003](#); [Hot et al. 2008](#); [McKendrick et al. 2004, 2005](#); [McKendrick, Sampson, Walland, & Badcock 2007](#)). En ce qui concerne l'impact d'une baisse de sensibilité au contraste sur les processus de plus haut niveau qui mènent à la reconnaissance visuelle, la question a été beaucoup moins explorée. Une exception notable est l'étude expérimentale de [Lenoble et al. \(2016\)](#). Pour rappel, ces auteurs ont observé que comparativement à des participants sains, les patients atteints de GPAO catégorisaient difficilement des petites images d'objets présentées en vision centrale lorsque le contraste

des images était diminué de moitié. Comme nous l'avons proposé plus haut, l'hypothèse d'un modèle FP-peri qui serait déficitaire dans le glaucome pourrait rendre compte de ce type de résultats. Dans cette première expérience réalisée auprès de patients GPAO, nous avons donc testé comment l'altération de la vision périphérique impacte la reconnaissance de stimuli complexes (images de scènes et visages) et faiblement contrastés en vision centrale.

Dans cette expérience, la manipulation critique était le niveau de traitement visuel requis pour effectuer la tâche, alors même que le stimulus restait constant. Pour ce faire, nous avons d'abord évalué les aspects de bas niveau du traitement visuel à l'aide d'une simple tâche de détection de photographies de scènes ou de visages humains. Les instructions imitaient celles de la périmétrie statique automatisée. Nous avons ensuite utilisé une tâche de catégorisation de ces mêmes stimuli, évaluant ainsi des aspects de plus haut niveau du traitement visuel. En testant les capacités des patients à détecter et catégoriser les mêmes stimuli, nous avons pu comparer directement les traitements visuels de bas et haut niveau. Nous avons considéré la différence de performances entre les deux tâches comme un indicateur du coût des processus visuels de haut niveau permettant d'accéder à la reconnaissance visuelle. Dans cette expérience, les patients GPAO ont été divisés en deux groupes sur la base de l'examen de la périmétrie automatisée : ceux dont le CV central était partiellement affecté et ceux dont le CV central était préservé. Nous avons comparé les performances des deux groupes de patients à celles d'un groupe contrôle, c'est-à-dire des participants appariés en âge qui n'avaient pas d'atteinte périmétrique.

Notre hypothèse était que le coût cognitif des processus menant à la reconnaissance visuelle soit plus important pour les patients que pour les participants contrôles. Nous avons donc fait une hypothèse d'interaction entre le Groupe (patients *vs* contrôles) et la Tâche, c'est-à-dire que nous nous attendions à ce que la différence de performances entre la tâche de détection et la tâche de catégorisation soit plus importante pour les patients que pour les contrôles. Comme nous nous attendions à observer cette interaction pour les deux groupes de patients (avec et sans atteinte du CV central), nous avons testé l'effet d'interaction entre le Groupe et la Tâche en considérant seulement le sous-groupe de patients avec une atteinte du CV central, puis en considérant seulement le sous-groupe de patients sans atteinte du CV central. Nous avons ensuite utilisé des tests de contraste pour comparer les groupes sur la tâche de détection d'une part, et sur la tâche de catégorisation d'autre part. Sur la base des résultats de la périmétrie statique automatisée, nous nous attendions à ce que seuls les patients présentant une atteinte du CV central aient des capacités modifiées pour les processus de bas niveau (tâche de détection) par rapport aux participants contrôles. Les performances des patients sans atteinte du CV central devaient être préservées pour cette tâche purement sensorielle. Par contre, nous nous attendions à ce qu'à la fois les patients avec un déficit du CV central et les patients sans déficit du CV central aient de moins bonnes performances que les participants contrôles à la tâche de catégorisation.

Par ailleurs, nous avons considéré l'impact du niveau de contraste de luminance des stimuli. Les stimuli étaient présentés à deux niveaux de contraste très faibles (contraste *Root-Mean-Square*, $RMS = 0.03$ et $RMS = 0.007$ pour des valeurs de luminance des pixels comprises entre 0 et 1)¹, mimant une situation de crépuscule ou plus généralement de faible éclairage. Nous nous attendions à ce que l'effet d'interaction entre le Groupe et la Tâche soit d'autant plus grand que le contraste est faible (i.e. $RMS = 0.007$).

2.1.1. Participants

Nous avons recruté des patients atteints de GPAO dans le cadre de leur suivi ophtalmologique au sein de la clinique d'ophtalmologie du Centre Hospitalier Universitaire Grenoble Alpes. Les patients étaient testés sur un seul oeil. L'acuité visuelle de l'oeil testé devait être au moins égale à 6.3/10 sur une échelle de Monoyer pour la vision de loin (ou 0.2 logMAR) et de 1.5 ou 2 sur une échelle de Parinaud pour la vision de près. Si les deux yeux correspondaient à ces critères, l'oeil testé était choisi en fonction des contraintes de la consultation ayant précédé l'inclusion (e.g., une dilatation pupillaire par collyre d'un des deux yeux entraînait l'exclusion de cet oeil), mais aussi afin d'équilibrer les deux groupes (avec ou sans atteinte du CV central). Vingt-deux patients (12 femmes, âge moyen : 67.5 ± 9 ans ; de 52 à 79 ans) atteints de GPAO ont été inclus dans l'expérience sur la base de ces critères. Le CV monoculaire des patients était évalué pour chaque oeil à l'aide d'un analyseur périmétrique Humphrey (Carl Zeiss, Meditec, Dublin, CA) avec une procédure SITA-Standard 24-2. Cet examen testait ainsi le seuil de sensibilité de 54 points du CV, séparés de 6 degrés d'angle visuel dans les 24 degrés centraux des côtés supérieur, temporal et inférieur, et dans les 30 degrés centraux du côté nasal. Si besoin, les patients réalisaient l'examen périmétrique avec une correction optique adaptée à leur réfraction. Nous avons évalué la sensibilité du CV central des patients sur la base de l'analyse statistique de la déviation individuelle de l'examen périmétrique. Une atteinte du CV central était définie par la présence d'au moins un point, parmi les quatre points testés dans un cercle central de 3 degrés de rayon, présentant une probabilité inférieure ou égale à 2% d'être normal. Les déficits constatés sur la base de ce critère nous ont permis de constituer deux groupes de patients (cf Figure 3.5) : 11 patients qui présentaient une atteinte du champ visuel central (glaucome-AC ; 5 femmes ; âge moyen : 68 ± 10 ans, min = 52, max = 79) et 11 patients qui ne présentaient pas d'atteinte du champ visuel central (glaucome-nAC ; 7 femmes : âge moyen : 67 ± 9 ans, min = 55, max = 79). L'examen périmétrique montrait une atteinte du CV périphérique pour tous les patients des deux groupes.

1. Lorsque nous avons publié ces résultats dans la revue PloS One, ces deux niveaux de contraste étaient labellisés 2.5% et 10%, exprimant le pourcentage de contraste restant par rapport au contraste de l'image non modifiée. Cependant, ce calcul peut être fait et interprété de plusieurs façons, et ces valeurs n'étaient pas nécessairement comparables à d'autres contrastes exprimés en pourcentage dans la littérature. Notamment, Lenoble et al. (2016) ont utilisé un contraste de Michelson. Les niveaux de contraste (50% et 100%) utilisés par ces auteurs sont donc difficilement comparables avec les niveaux de contraste de l'Expérience 6. Pour plus de clarté, nous avons donc choisi d'exprimer seulement les valeurs de contrastes RMS (correspondant à l'écart-type de la luminance sur l'image).

Nous avons également recruté 25 participants volontaires sans glaucome, appariés en âge aux groupes de patients (14 femmes; âge moyen : 68 ± 6 ans; min = 54, max = 83) afin de constituer le groupe contrôle. Ces participants ont également réalisé une évaluation périmétrique Humphrey (SITA-Standard 24-2) sur chaque oeil pour s'assurer qu'il n'y avait pas de déficit de sensibilité rétinienne dans leur CV. Nous les avons ensuite testé sur un des deux yeux. Sur cette oeil, l'acuité avec correction de la réfraction était de 10/10 sur une échelle de Monoyer et de 2 sur une échelle de Parinaud pour tous les participants.

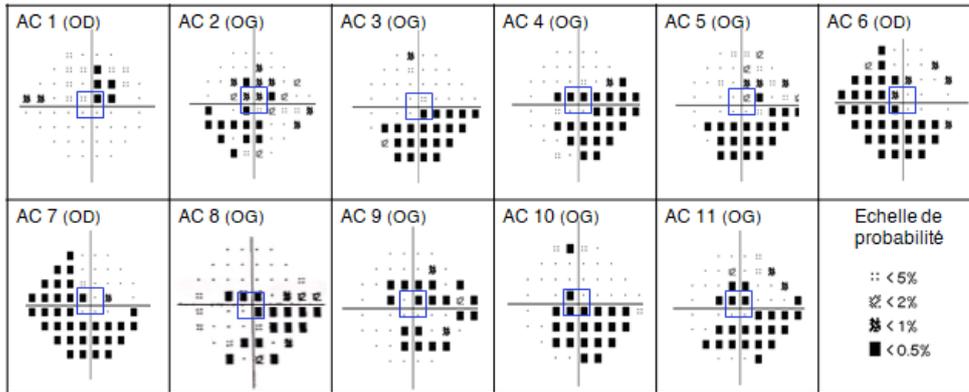
Un examen ophtalmologique complet (acuité visuelle, tonométrie de Goldmann, segment antérieur, fond de l'oeil) a été réalisé pour chaque participant des trois groupes (contrôle, glaucome-AC, glaucome-nAC). Les participants ne devaient pas présenter d'autres pathologies oculaires (notamment une dégénérescence maculaire liée à l'âge ou une cataracte non opérée), ni de pathologies neurologiques ou psychiatriques. Les participants dont la cataracte avait été opérée étaient inclus seulement s'il n'y avait pas eu de complication à l'issue de l'opération. Dans tous les cas le cristallin était clair, chez les patients atteints de glaucome comme chez les participants contrôles. Nous avons également rapidement évalué les fonctions cognitives générales à l'aide d'un MMSE (*Mini Mental State Examination*). Le score au MMSE devait être d'au moins 24/30 pour tous les participants. Tous ont donné leur consentement libre et éclairé avant de participer à l'expérience qui a été menée en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine, et approuvée par un comité d'éthique (CERNI², COMUE Université Grenoble Alpes : IRB00010290-2016-09-13-1).

2.1.2. Stimuli et procédure

Les participants réalisaient deux sessions expérimentales, la session SCENE et la session VISAGE. Les stimuli utilisés pour la session SCENE sont des photographies de scènes en noir et blanc (256 niveaux de gris) issues de la base d'images ReCor (Musel et al. 2013; Peyrin 2018). Nous avons sélectionné 20 scènes (192×144 pixels) appartenant à deux catégories sémantiques : 10 scènes d'intérieur (cuisines, salons, ...) et 10 scènes d'extérieur (rues, bâtiments, maisons, ...). Les spectres d'amplitude moyens (i.e. la distribution de l'énergie sur les différentes fréquences spatiales et orientations) des deux catégories étaient similaires. Ainsi, il était difficile de se baser sur les régularités statistiques des caractéristiques physiques propres à chaque catégorie pour discriminer les deux catégories. En d'autres termes, la reconnaissance d'un environnement d'intérieur ou d'extérieur était nécessaire pour réaliser la tâche de catégorisation, puisqu'elle ne pouvait se baser seulement sur ces indices physiques. Les stimuli de la session VISAGE sont des photographies en noir et blanc (256 niveaux de gris) issues de précédentes

2. Le CERNI (Comité d'Éthique pour les Recherches Non Interventionnelles) est un comité d'éthique local qui a évalué les recherches en santé non interventionnelles menées avant l'entrée en vigueur de la loi n°2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (dite loi Jardé) depuis la publication des décrets d'application le 16 novembre 2016.

Patients avec une atteinte du champ visuel central (AC)



Patients sans atteinte du champ visuel central (nAC)

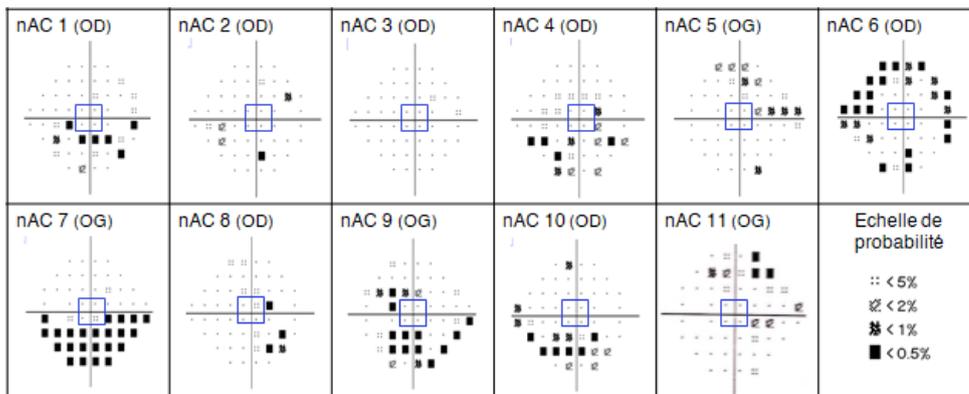


Figure 3.5: *Probability plots* des déviations individuelles monoculaires des patients GPAO inclus dans l'Expérience 6, obtenus avec une périmétrie statique automatisée de Humphrey (procédure SITA-Standard 24-2). Un déficit dans le champ visuel central était défini par au moins un des quatre points centraux (encadrés en bleu) présentant une probabilité de 2% ou moins d'être normal. L'oeil testé est indiqué pour chaque patient (OG : oeil gauche ; OD : oeil droit). AC : Patients avec une atteinte du champ visuel central ; nAC : Patients sans atteinte du champ visuel central.

2. Expérience 6 : Étude de la reconnaissance de scènes et de visages faiblement contrastés dans la vision centrale de patients atteints d'un glaucome

recherches menées par les membres du LPNC (Guyader et al. 2017; Kauffmann et al. 2019). Nous avons sélectionné 10 visages humains, cinq animaux et cinq véhicules (192×192 pixels).

L'expérience était programmée avec E-Prime 2. Les stimuli étaient affichés sur un écran de 30" d'une résolution de 2560×1600 pixels, avec un taux de rafraîchissement de 60Hz. Les participants étaient assis dans l'obscurité à une distance de 38 cm de l'écran. À cette distance, les stimuli respectaient une taille de 6° d'angle visuel de largeur. Ainsi, la région du CV stimulée correspondait à la région contenant les quatre points centraux de l'examen de la périmétrie automatisée utilisée pour déterminer si les patients avaient un déficit du CV central. Une correction de la réfraction adaptée au participant et à la distance de l'écran était utilisée (soit avec leurs propres lunettes, soit avec des verres d'essai). Pour maintenir une position centrée et la distance à l'écran, la tête des participants reposait sur une mentonnière positionnée dans l'axe sagittal médian.

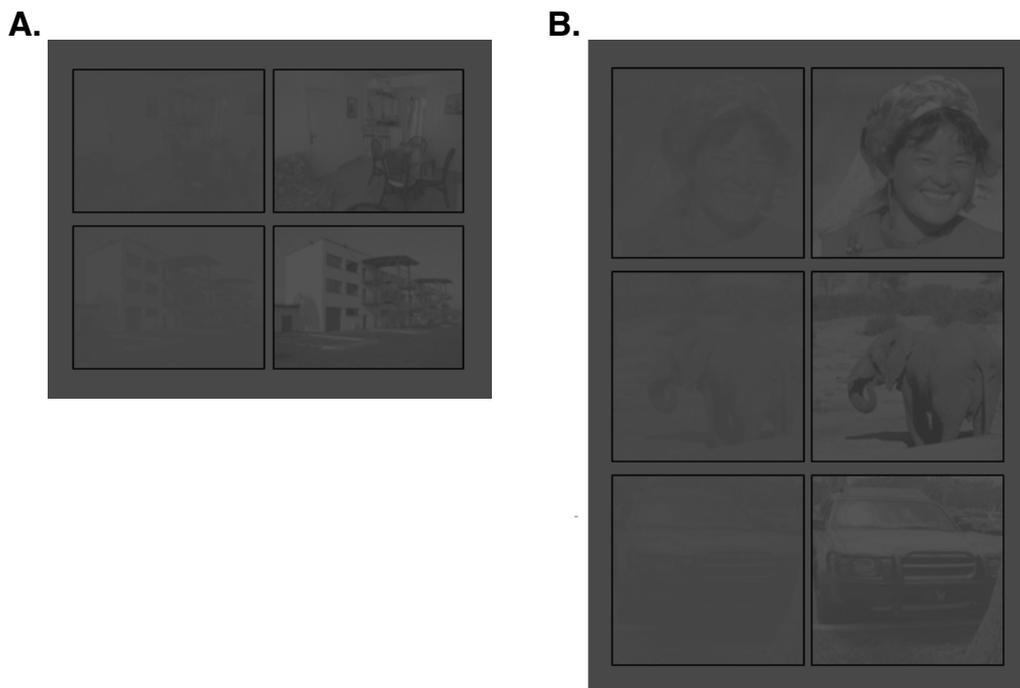


Figure 3.6: Exemples de stimuli utilisés dans l'Expérience 6. **A.** Session expérimentale SCENE, **B.** Session expérimentale VISAGES. Les images étaient présentées à deux niveaux de contrastes faibles (RMS de 0.007 et 0.03).

Nous avons mesuré la fonction gamma de l'écran (i.e. la luminance réelle de l'écran pour différents niveaux de gris) en utilisant un outil de calibration (Spyder5Elite). Nous avons ensuite corrigé la luminance des pixels de chaque stimulus pour linéariser indirectement la fonction d'affichage de l'écran. Ainsi, la manipulation du contraste de luminance des images avait un effet proportionnel sur l'ensemble des valeurs de luminance, depuis les bas niveaux de luminance (zones sombres) jusqu'aux plus hauts niveaux (zones claires).

La manipulation du contraste a été réalisée dans l'environnement MATLAB. Nous avons créé deux versions de chaque image, en manipulant le contraste RMS : une version avec un RMS de 0.007 et une version avec un RMS de 0.03 (pour des valeurs de luminance allant de 0 à 1 ; cf. Figure 3.6). La luminance moyenne de toutes les images, ainsi que la luminance du fond gris uniforme sur lequel les stimuli étaient affichés, était de 72 pour une échelle de 256 niveaux de gris (soit 0.28 pour des valeurs de luminance comprises entre 0 et 1 ; ou approximativement 10 candela/m²). Cette valeur correspondait à la luminance du fond de la coupole de la périmétrie automatisée de Humphrey (luminance de 31.5 apostilbs, correspondant à environ 72 niveaux de gris ou 10 candela/m²). Chaque session expérimentale (session SCENE et session VISAGE) comprenait deux blocs. Dans un bloc, les participants réalisaient une tâche de détection et dans l'autre une tâche de catégorisation.

Lors du bloc de détection de la session SCENE, les participants avaient pour tâche d'appuyer sur la barre espace d'un clavier d'ordinateur chaque fois qu'ils considéraient qu'une image apparaissait à l'écran, et de ne pas appuyer sinon (paradigme go/no-go). Le bloc de détection comprenait 20 essais go et 20 essais no-go. Dans les essais go (10 essais pour chaque niveau de contraste), une scène était présentée sur le fond gris uniforme. Dans les essais no-go, aucune image n'était présentée, seul le fond gris uniforme de l'écran. L'attribution de la catégorie des scènes présentées lors des essais go était contrebalancée entre les participants de chaque groupe (contrôle, glaucome-AC, glaucome-nAC). Ainsi, les stimuli des essais go étaient des scènes d'intérieur pour la moitié des participants de chaque groupe, et des scènes d'extérieur pour l'autre moitié. On précisait explicitement aux participants qu'ils ne devaient pas essayer de reconnaître ce qu'il y avait sur l'image. Ils devaient répondre aussi vite que possible, dès qu'ils percevaient la présence d'une image. Lors du bloc de catégorisation de la session SCENE, les participants avaient pour tâche d'appuyer sur la barre espace lorsque l'image appartenait à la catégorie cible, et de retenir leur réponse lorsqu'elle appartenait à la catégorie distractive. Le bloc de catégorisation comprenait 20 essais go comprenant une scène de la catégorie cible (10 pour chaque niveau de contraste) et 20 essais no-go comprenant une scène de la catégorie distractive (10 pour chaque niveau de contraste). Pour chaque participant, nous avons veillé à ce que la catégorie cible corresponde à la catégorie présentée lors des essais go de la tâche de détection. Ainsi, un participant donné devait détecter et catégoriser soit des scènes d'extérieur, soit des scènes d'intérieur.

Lors du bloc de détection de la session VISAGE, la tâche était la même que pour la session SCENE. Le bloc de détection comprenait 20 essais go et 20 essais no-go. Dans les essais go, les stimuli présentés étaient toujours des visages (10 essais pour chaque niveau de contraste). Dans les essais no-go, aucune image n'était présentée. De la même façon que pour la session SCENE, on précisait explicitement aux participants qu'ils ne devaient pas essayer de reconnaître ce qu'il y avait sur l'image. Ils devaient répondre aussi vite que possible, dès qu'ils percevaient la présence d'une image. Lors du bloc de catégorisation, les participants avaient pour tâche d'appuyer sur la barre d'espace

lorsque l'image appartenait à la catégorie cible, et de retenir leur réponse lorsqu'elle appartenait à une catégorie distractive. La catégorie cible était toujours les visages, et les distracteurs étaient toujours les animaux et les véhicules. La tâche de catégorisation comprenait également 20 essais go avec une image de visage (10 pour chaque niveau de contraste) et 20 essais no-go avec des images de véhicules et d'animaux (10 de chaque catégorie distractive, et 10 pour chaque niveau de contraste). Les participants devaient répondre aussi vite que possible, dès qu'ils considéraient qu'un visage apparaissait sur l'écran. Là encore, les participants commençaient toujours la session par le bloc de détection.

Dans chaque session, les stimuli des essais go étaient exactement les mêmes dans les blocs détection et catégorisation (des scènes d'intérieur ou des scènes d'extérieur pour la session SCENE et des visages pour la session VISAGE) et seules les demandes de la tâche changeaient. À chaque essai, un point de fixation apparaissait au centre de l'écran pendant 500 ms, puis un stimulus (ou seulement le fond gris uniforme pour les essais no-go des blocs de détection) était affiché pendant 100 ms sur le fond gris. Le temps de présentation était choisi afin de permettre un temps de traitement suffisamment long et confortable pour les participants, tout en minimisant la possibilité de faire des saccades oculaires. Le point de fixation était à nouveau affiché après la présentation du stimulus (ou du fond gris uniforme) pendant 1900 ms afin que les participants puissent répondre tout en maintenant leur regard au centre de l'écran. Les participants pouvaient donner leur réponse avec la main de leur choix. Les consignes étaient affichées à l'écran avant chaque bloc expérimental, mais elles étaient également détaillées à l'oral par l'expérimentatrice. Avant chaque bloc expérimental, les participants réalisaient un court bloc d'entraînement comprenant 12 essais, avec des stimuli différents de ceux utilisés pendant l'expérience. L'ordre des sessions SCENE et VISAGE était contrebalancé entre les participants au sein de chaque groupe. Dans chaque session expérimentale, les participants réalisaient toujours le bloc de détection avant le bloc de catégorisation, afin d'éviter la contamination de processus de catégorisation sur la tâche de détection. L'ordre de présentation des conditions de contraste était randomisé à l'intérieur de chaque bloc expérimental. La réponse du participant était enregistrée pour chaque essai (soit une réponse go, soit une réponse no-go) et les temps de réponse étaient enregistrés pour les réponses go.

2.1.3. Analyse des données

Nous avons analysé les données à l'aide de la Théorie de Détection du Signal en étiquetant chaque réponse comme étant une détection correcte (de l'image dans la tâche de détection ou de la catégorie cible dans la tâche de catégorisation), une omission, une fausse alarme ou un rejet correct. L'indice de sensibilité d' était ensuite calculé pour chaque participant et chaque condition expérimentale sur la base des proportions de chaque type de réponse. Les temps de réponses correctes moyens, qui nous intéressaient initialement, n'ont pas pu être calculés à cause de proportions extrêmes d'omissions pour

plusieurs patients, surtout pour le bloc de catégorisation lorsque le contraste RMS était de 0.007.

Pour chaque session expérimentale (session SCENE et session VISAGE), nous avons ensuite analysé les d' à l'aide d'une analyse de variance à mesures répétées. Nous avons entré dans cette analyse le Groupe comme facteur inter-sujet (contrôle, glaucome-AC, glaucome-nAC), et la Tâche (détection, catégorisation) et le niveau de Contraste (RMS = 0.03, RMS = 0.007) comme facteurs intra-sujets. Les tailles d'effets sont reportés en eta-carré partiels (η^2), qui représentent la proportion de variance pouvant être attribuée à chacun des effets testés. La significativité des effets a été testée à un seuil alpha de 0.05. Afin de rester dans la logique d'analyse adoptée dans cette thèse, nous avons évité de reporter ici une analyse exhaustive du plan factoriel (test de tous les effets principaux, effets d'interaction et effets simples, cf. Chapitre 2, partie 1). Nous reportons ici uniquement les effets d'intérêt.

2.2. Résultats

2.2.1. Session expérimentale 'SCENE'

Les moyennes et écart-types des d' pour chaque condition expérimentale et chaque groupe sont reportés dans le Tableau 3.2.

Nous avons tout d'abord testé si la différence de sensibilité (d') entre les tâches de détection et de catégorisation variait en fonction du groupe (interaction Groupe*Tâche). Nos analyses montrent un effet d'interaction omnibus entre le Groupe et la Tâche ($F(2,44) = 8.27$, $p < .001$, $\eta^2 = .27$; cf. Figure 3.7-A), indiquant que la magnitude de la différence de d' entre les tâches de détection et de discrimination dépend du groupe (contrôle, glaucome-AC, glaucome-nAC). Cet effet d'interaction est significatif à la fois lorsqu'on ne considère que les groupes contrôle et glaucome-AC ($F(1,44) = 15.73$, $p < .001$, $\eta^2 = .32$) et lorsqu'on ne considère que les groupes contrôle et glaucome-nAC ($F(1,44) = 4.26$, $p = .045$, $\eta^2 = .12$). Autrement dit, la différence de d' entre les deux tâches est plus importante pour les patients glaucomateux que pour les participants contrôles, que les patients aient une atteinte du CV central ou non. La taille d'effet est cependant plus grande pour les patients glaucome-AC que pour les patients glaucome-nAC.

Tableau 3.2: Résultats de l'Expérience 6 (session SCENE). Indice d' en fonction du groupe (Contrôle, glaucome-nAC : sans atteinte du champ visuel central, glaucome-AC : avec atteinte du champ visuel central), de la tâche (Détection, Catégorisation), et du niveau de contraste (RMS-0.007; RMS-0.03)

	contrôle		glaucome-nAC		glaucome-AC	
	Détection	Catégorisation	Détection	Catégorisation	Détection	Catégorisation
RMS = 0.007	3.50 ±0.64	2.21 ±0.84	3.13 ±0.80	1.21 ±1.19	2.41 ±1.38	0.59 ±1.04
RMS = 0.03	3.91 ±0.28	3.35 ±0.69	3.68 ±0.45	2.71 ±0.93	3.71 ±0.52	1.70 ±0.82

2. Expérience 6 : Étude de la reconnaissance de scènes et de visages faiblement contrastés dans la vision centrale de patients atteints d'un glaucome

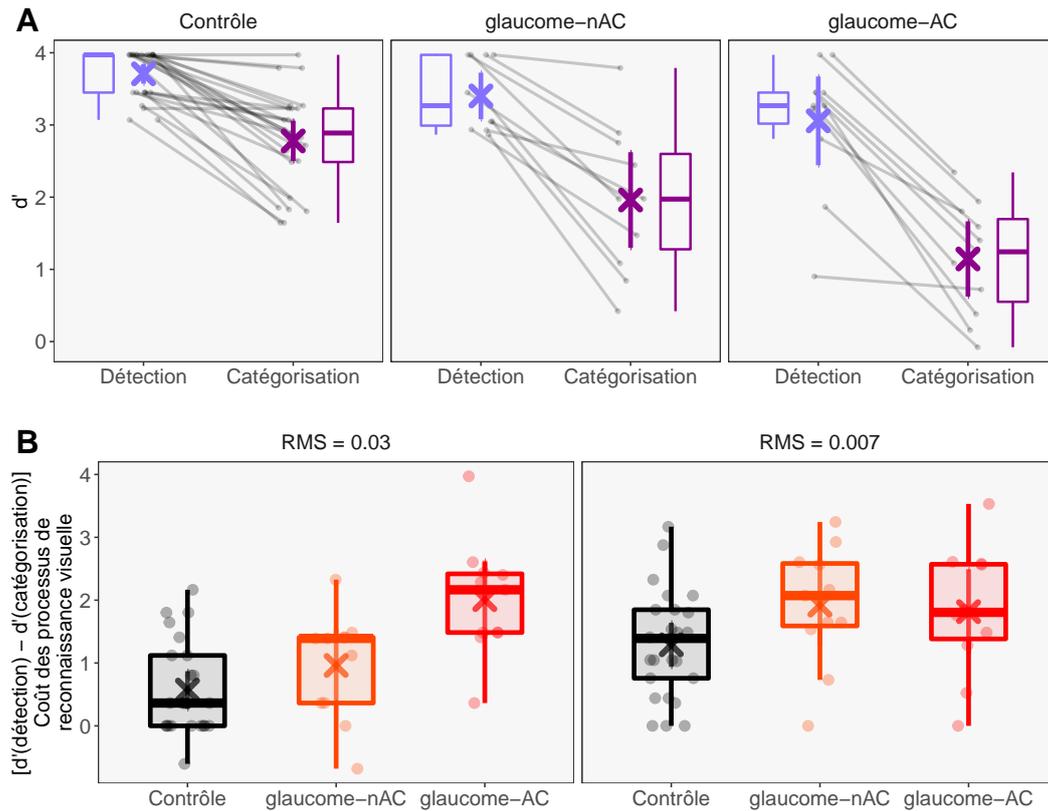


Figure 3.7: Résultats de l'Expérience 6 (session SCENE). **A.** Indice d' en fonction du Groupe (Contrôle, glaucome-nAC : sans atteinte du champ visuel central, glaucome-AC : avec atteinte du champ visuel central) et de la tâche (Détection, Catégorisation). **B.** Différence de d' entre la tâche de détection et la tâche de catégorisation [d' (détection) - d' (catégorisation)] en fonction du Groupe (Contrôle, glaucome-nAC, glaucome-AC) et du niveau de contraste (RMS-0.03 et RMS-0.007). La différence de d' entre les deux tâches représente le coût cognitif des processus menant à la reconnaissance visuelle.

Nous avons ensuite comparé les différences de d' entre les trois groupes pour chaque tâche séparément à l'aide de tests de contrastes. Pour la tâche de détection, la sensibilité des patients glaucome-AC était moins bonne que celle des participants contrôles ($F(1,44) = 10.74$, $p = .002$, $\eta^2 = .23$) alors qu'il n'y a pas de différence significative entre la sensibilité des patients glaucome-nAC et celle des participants contrôles ($F(1,44) = 2.38$, $p = .130$). Pour la tâche de catégorisation, la sensibilité des patients glaucome-AC était moins bonne que celle des participants contrôles ($F(1,44) = 33.71$, $p < .001$, $\eta^2 = .54$). Similairement, la sensibilité des patients glaucome-nAC à cette tâche de catégorisation était également moins bonne que celle des participants contrôles, bien que l'effet soit de plus petite taille ($F(1,44) = 8.45$, $p = .006$, $\eta^2 = .20$).

Nous avons ensuite considéré comment le niveau de contraste de luminance pouvait moduler l'interaction entre le Groupe et la Tâche. Pour clarifier cet effet de double interaction, la Figure 3.7-B montre la différence de d' entre la tâche de détection et la

tâche de catégorisation [d' (détection) - d' (catégorisation)]. Cette différence correspond au coût cognitif de la catégorisation (i.e. des processus menant à la reconnaissance visuelle) pour chaque groupe et pour chaque niveau de contraste. Nous avons d'abord testé l'interaction Groupe*Tâche*Contraste en considérant seulement le groupe contrôle et le groupe glaucome-AC. Cette interaction est significative. Le contraste n'avait pas le même effet sur la différence entre les tâches de détection et de catégorisation dans ces deux groupes ($F(1,44) = 5.83$, $p < 0.05$, $\eta^2 = .15$). Le coût de la catégorisation est plus fort pour les patients AC que pour les participants contrôles lorsque le contraste RMS est de 0.03 ($F(1,44) = 24.39$, $p < .001$), mais pas lorsque le contraste RMS est de 0.007 ($F(2,44) = 2.55$, $p = .118$). Ce dernier résultat est surprenant, mais peut potentiellement s'expliquer par un effet plancher : la sensibilité pour la tâche de catégorisation lorsque le RMS est de 0.007 est basse, même, dans une moindre mesure, pour le groupe contrôle (cf. Tableau 3.2). Nous avons ensuite testé l'interaction Groupe*Tâche*Contraste en considérant seulement le groupe contrôle et le groupe glaucome-nAC. L'interaction n'est pas significative ($F(1,44) < 1$).

2.2.2. Session expérimentale 'VISAGE'

Les moyennes et écart-types des d' pour chaque condition expérimentale et chaque groupe sont reportés dans le Tableau 3.3.

Nous avons tout d'abord comparé la différence d' entre les tâches de détection et de catégorisation en fonction du groupe (interaction Groupe*Tâche). Nos analyses montrent un effet d'interaction omnibus entre le Groupe et la Tâche ($F(2,44) = 5.31$, $p = 0.009$, $\eta^2 = 0.19$; cf. Figure 3.8-A), indiquant que la magnitude de la différence de d' entre les tâches de détection et de discrimination dépend du groupe (contrôle, glaucome-AC, glaucome-nAC). Cet effet d'interaction est significatif lorsqu'on ne considère que le groupe contrôle et le groupe glaucome-AC ($F(1,44) = 8.62$, $p = 0.005$, $\eta^2 = 0.19$). Par contre, contrairement à la session expérimentale SCENE, cet effet d'interaction n'est pas significatif lorsqu'on ne considère que le groupe contrôle et le groupe glaucome-nAC ($F(1,44) < 1$). En d'autres termes, la différence de d' entre les deux tâches est plus importante pour les patients qui ont une atteinte du CV central que pour les participants contrôles; en revanche, la différence de d' entre les deux tâches pour les patients qui n'ont pas d'atteinte du CV central n'est pas significativement différente de celle des participants contrôles, ces deux groupes ayant une bonne sensibilité tant pour détecter que pour catégoriser les visages.

Nous avons ensuite comparé les d' des trois groupes pour chaque tâche séparément. Pour la tâche de détection, la sensibilité des patients AC est moins bonne que celle des participants contrôles ($F(1,44) = 16.24$, $p < 0.001$, $\eta^2 = 0.33$), alors qu'il n'y a pas de différence significative entre la sensibilité des patients nAC et celle des participants contrôles ($F(1,44) = 2.63$, $p = 0.112$). Pour la tâche de catégorisation, la sensibilité des patients AC est également moins bonne que celle des participants contrôles ($F(1,44) = 26.80$, $p < 0.001$, $\eta^2 = 0.43$). Par contre, dans cette session, les patients nAC ont une

2. *Expérience 6 : Étude de la reconnaissance de scènes et de visages faiblement contrastés dans la vision centrale de patients atteints d'un glaucome*

aussi bonne sensibilité que celle des participants contrôles dans la tâche de catégorisation de visages ($F(1,44) < 1$).

Tableau 3.3: Résultats de l'Expérience 6 (session VISAGE). Indice d' en fonction du groupe (Contrôle, glaucome-nAC : sans atteinte du champ visuel central, glaucome-AC : avec atteinte du champ visuel central), de la tâche (Détection, Catégorisation), et du niveau de contraste (RMS-0.007; RMS-0.03)

	contrôle		glaucome-nAC		glaucome-AC	
	Détection	Catégorisation	Détection	Catégorisation	Détection	Catégorisation
RMS = 0.007	3.76 ± 0.39	3.55 ± 0.60	3.29 ± 0.81	3.05 ± 0.99	2.86 ± 1.25	1.58 ± 1.25
RMS = 0.03	3.96 ± 0.07	3.65 ± 0.49	3.97 ± 0.00	3.87 ± 0.17	3.71 ± 0.45	3.33 ± 0.88

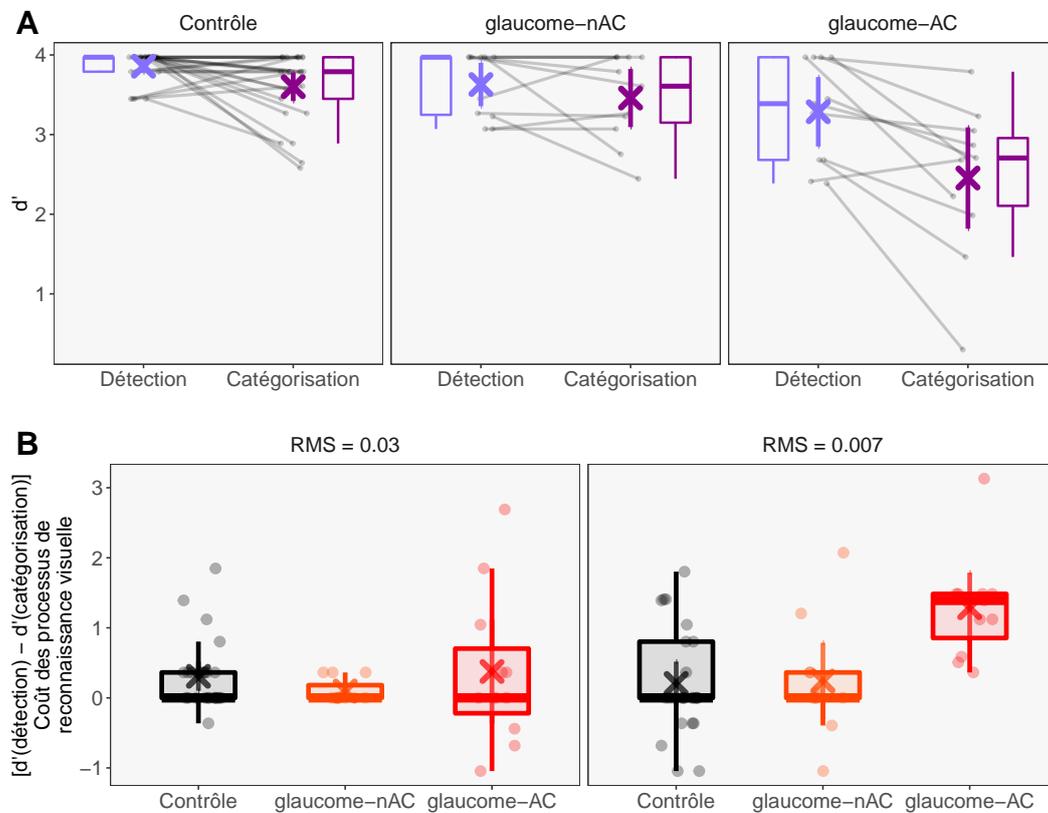


Figure 3.8: Résultats de l'Expérience 6 (session VISAGE). **A.** Indice d' en fonction du Groupe (Contrôle, glaucome-nAC : sans atteinte du champ visuel central, glaucome-AC : avec atteinte du champ visuel central) et de la tâche (Détection, Catégorisation). **B.** Différence de d' entre la tâche de détection et la tâche de catégorisation [$d'(\text{détection}) - d'(\text{catégorisation})$] en fonction du Groupe (Contrôle, glaucome-nAC, glaucome-AC) et du niveau de contraste (RMS = 0.03 et RMS = 0.007). La différence de d' entre les deux tâches représente le coût cognitif des processus menant à la reconnaissance visuelle.

Nous avons ensuite considéré comment le niveau de contraste de luminance pouvait moduler l'interaction entre le Groupe et la Tâche (Figure 3.8-B). Nous avons d'abord

testé l'interaction Groupe*Tâche*Contraste en considérant seulement le groupe contrôle et le groupe glaucome-AC. Cet effet d'interaction est significatif : le contraste n'avait pas le même effet sur la différence entre les tâches de détection et de discrimination dans ces deux groupes ($F(1,44) = 9.24, p = .004, \eta^2 = .19$). Le coût de la catégorisation pour les patients AC n'est pas significativement différent de celui des participants contrôles lorsque le contraste RMS est de 0.03 ($F(1,44) < 1$), mais il est plus grand que celui des participants contrôles lorsque le contraste RMS est de 0.007 ($F(2,44) = 15.19, p < .001$). Nous avons ensuite testé l'interaction Groupe*Tâche*Contraste en considérant seulement le groupe contrôle et le groupe glaucome-nAC. Cette interaction n'est pas significative ($F(1,44) < 1$) suggérant de nouveau que le niveau de contraste ne modulait pas le coût des processus de haut niveau chez ces patients.

2.3. Discussion

Cette étude avait pour but d'étudier l'impact du glaucome sur la reconnaissance, en vision centrale, de stimuli visuels complexes à faible niveau de contraste. Nos analyses ont montré que les résultats diffèrent en fonction de la présence d'une atteinte périmétrique ou non dans le CV central (i.e. patients glaucome-AC et patients glaucome-nAC). Dans un premier temps, on constate que les patients AC présentent un déficit pour détecter et pour catégoriser des stimuli faiblement contrastés (scènes et visages), relativement aux participants contrôles appariés en âge. Ce résultat est prévisible sur la seule base de l'examen périmétrique, qui montre une atteinte du CV central chez ces patients. Les données de ces patients qui ont une atteinte du CV central viennent appuyer celles d'études précédentes qui montraient que des patients qui ont une atteinte périmétrique dans les 10° centraux du CV ont des difficultés pour reconnaître les visages (Glen et al. 2012, 2013). Glen et al. (2013) ont par ailleurs montré que les performances de ces patients sur une tâche de reconnaissance de visages étaient meilleures lorsque le nombre de fixations était moins important et lorsque l'amplitude des saccades était plus grande. Le déficit pourrait donc être compensé grâce à l'engagement de patterns spécifiques de mouvements oculaires (des saccades moins nombreuses mais plus amples). Dans notre étude, les stimuli étaient présentés pendant 100 ms seulement, ce qui empêchait les patients avec une atteinte du CV central de s'appuyer sur de telles stratégies. À noter que les difficultés de ces patients dans la tâche de catégorisation de visages interviennent seulement lorsque le contraste RMS des stimuli est extrêmement bas (0.007). Par ailleurs, nous avons observé un effet d'interaction entre le groupe (participants contrôles - patients glaucome-AC) et la tâche (détection - catégorisation) quels que soient les stimuli. La différence de performances entre la détection et la catégorisation était plus élevée chez ces patients AC que chez les participants contrôles, suggérant un coût de traitement supplémentaire pour réaliser la tâche de catégorisation pour ces patients. Deux cas de figure peuvent expliquer cet effet d'interaction :

- (i) Dans le premier cas de figure, le traitement des entrées sensorielles disponibles

en vision centrale est impacté pour les patients AC. Le coût de traitement ([détection - catégorisation]) plus important pour les patients que pour les contrôles pourrait alors s'expliquer par le fait que chaque tâche ne nécessite pas la même quantité d'information pour être correctement réalisée. En effet, dans la tâche de catégorisation, il est préférable de percevoir la totalité du stimulus visuel pour aboutir à la reconnaissance. Par contre, dans la tâche de détection, l'intégralité du stimulus visuel n'est pas nécessaire pour décider qu'un stimulus est présent ou non. La perception d'une petite partie de l'image suffit à sa détection. Dans ce cas figure, les déficits portent seulement sur des traitements ascendants.

- **(ii)** Dans le second cas de figure, ce sont les mécanismes de prédiction des patients AC qui seraient impactés. Le coût de traitement ([détection - catégorisation]) plus important pour les patients que pour les contrôles s'expliquerait par un dysfonctionnement du système prédictif causé par l'altération de la rétine périphérique. Dans ce cas figure, les déficits portent à la fois sur des traitements ascendants et descendants.

À noter que ces deux cas de figure ne sont pas forcément incompatibles et peuvent s'exprimer de façon concomitante. Un moyen de dissocier ces deux interprétations serait de s'assurer que l'entrée rétinienne est la même entre le groupe de patients et le groupe contrôle. On pourrait alors présenter aux participants contrôles des stimuli qui considéreraient les scotomes de chaque patient en ajoutant, par exemple, des tâches noires. Malheureusement, ce type de manipulation expérimentale est très difficile à appliquer en pratique et ne correspond pas forcément à la réalité phénoménologique du patient. En effet, l'identification d'un scotome grâce à l'examen ophtalmologique (périmétrie) n'implique pas une perception en tout ou rien. Le groupe de patients nAC nous permet en revanche d'étudier un déficit de catégorisation en nous affranchissant du problème du déficit du traitement des entrées sensorielles en vision centrale.

Concernant à présent ces patients nAC, les résultats diffèrent entre les sessions. Pour la session SCENE, nos analyses montrent que les patients nAC ont des performances de détection préservées, mais que leurs performances pour la tâche de catégorisation de scènes sont moins bonnes que celles des participants contrôles. Comme pour les patients AC, nous avons également observé un effet d'interaction entre le groupe et la tâche, de telle sorte que la différence de d' entre la détection et la catégorisation était plus élevée chez les patients nAC que chez les participants contrôles. Ce résultat suggère à nouveau que le coût des processus menant à la catégorisation est plus élevé chez les patients atteints de glaucome que chez les participants contrôles. Chez ces patients qui n'ont pas d'atteinte du CV central, le coût de traitement supplémentaire pour réaliser la tâche de catégorisation résulterait alors d'un dysfonctionnement du système prédictif (i.e. l'hypothèse **(ii)** décrite plus haut), puisqu'il ne peut pas s'expliquer par la baisse de sensibilité rétinienne en vision centrale (i.e. l'hypothèse **(i)** décrite plus haut). À noter toutefois que les patients nAC ont montré des performances similaires à celle des

participants contrôles pour la détection des visages mais également pour leur catégorisation. Les processus de reconnaissance des visages semblent donc moins impactés, puisque seuls les patients AC qui ont une atteinte du CV central ont un déficit pour catégoriser les visages, et que même chez ces patients, ce déficit se manifestait seulement pour les stimuli les moins contrastés (i.e. pour un RMS de 0.007, mais pas pour un RMS de 0.03). La divergence de résultats entre les sessions SCENE et VISAGE pourrait s'expliquer par les caractéristiques du traitement visuel de ces deux catégories. Comme nous l'avons vu dans le Chapitre 1 (sous-partie 3.1), la perception du *gist* des scènes est difficilement dissociable de la vision périphérique, une idée qui est notamment corroborée par les études montrant que la reconnaissance du *gist* est meilleure lorsque l'information centrale est masquée que lorsque l'information périphérique est masquée (Larson & Loschky 2009). Au contraire, les visages sont des stimuli qui recrutent plutôt la vision centrale. L'étude des mouvements oculaires lors de l'exploration de scènes visuelles montre par exemple que les individus dirigent automatiquement leur regard vers les visages dans la scène afin qu'ils soient traités en vision centrale (Coutrot & Guyader 2014; Foulsham, Cheng, Tracy, Henrich, & Kingstone 2010; Judd, Ehinger, Durand, & Torralba 2009). Au niveau neurofonctionnel, les régions visuelles de haut niveau dans le cortex occipito-temporal ventral qui sont spécifiquement impliquées dans le traitement des scènes et des visages présentent une forme d'organisation rétinotopique "secondaire" qui se traduit par une organisation de type centre-périphérie (Arcaro, McMains, Singer, & Kastner 2009; Baldassano, Fei-Fei, & Beck 2016; Hasson, Levy, Behrmann, Hendler, & Malach 2002; Levy, Hasson, Avidan, Hendler, & Malach 2001; Malach, Levy, & Hasson 2002). Par exemple, une étude en IRMf de Levy et al. (2001) a mis en évidence que des régions du cortex occipito-temporal activées par des stimuli de type scène (*vs* des stimuli de type visage) sont également activées par des stimulations dans le CV périphérique (*vs* des stimulations du CV central). Inversement, les régions activées par des stimuli de type visages (*vs* scènes) sont particulièrement activées par des stimulations dans le CV central (*vs* des stimulations du CV périphérique). Si un tel biais d'excentricité existe, la perception des visages serait davantage associée à la vision centrale et serait donc moins impactée par une perte de vision périphérique. Cette interprétation des résultats est malheureusement difficilement compatible avec notre hypothèse d'un dysfonctionnement du système prédictif dans le glaucome. En effet, dans le cadre du modèle FP-péri que nous proposons, si le système de gestion des prédictions fonctionne moins bien chez les personnes atteintes de glaucome, la reconnaissance visuelle devrait être perturbée quel que soit le type de stimulus. Il semble peu parcimonieux d'envisager l'existence d'un module prédictif spécifique au traitement des visages, qui lui ne serait pas impacté par le glaucome. Alternativement, la divergence des résultats obtenus entre les sessions SCENE et VISAGE pourrait résulter d'une différence de difficulté entre les deux tâches de catégorisation. En effet, le niveau de catégorisation pour la session SCENE (discriminer des scènes d'intérieur parmi des scènes d'extérieur, ou vice-versa) est plus bas que celui pour la session VISAGE (discriminer des visages parmi d'autres stimuli). D'ailleurs, les scènes d'intérieur partagent plus de caractéristiques visuelles avec

les scènes d'extérieur (notamment parce que les spectres d'amplitude moyens de ces deux catégories étaient contrôlés pour être similaires) que les visages avec les véhicules et les animaux. La reconnaissance de scènes parmi d'autres scènes peut donc être plus difficile que la reconnaissance de visages parmi des stimuli d'autres catégories, et ce d'autant plus que le contraste est très bas. La session SCENE serait donc plus adaptée que la session VISAGE pour mettre en évidence un dysfonctionnement du système prédictif.

Ainsi, les résultats de l'Expérience 6, et en particulier le pattern observé chez les patients nAC, suggèrent que le glaucome entraînerait un déficit visuel de haut niveau pour reconnaître des scènes dans le CV central qui ne peut pas être prédit par la seule baisse de sensibilité rétinienne. En ce sens, ces résultats sont donc compatibles avec la proposition théorique d'une altération du système de reconnaissance visuelle prédictif causée par l'altération du traitement des inputs provenant de la rétine périphérique (cf. Figure 1.21 dans le Chapitre 1), et permettent ainsi d'étayer l'hypothèse du modèle FP-peri. Cela dit, il faut souligner qu'il est possible que l'examen périmétrique n'était pas le plus approprié pour mettre en évidence une atteinte du CV central chez les patients nAC. En effet, la sensibilité rétinienne dans le CV a été mesurée grâce à une périmétrie statique automatisée de Humphrey, avec une procédure SITA-Standard 24-2. Dans la pratique clinique, si cette procédure révèle une mauvaise sensibilité sur un des points centraux ou paracentraux, on réalise parfois une évaluation plus spécifique du champ visuel central avec la procédure SITA-Standard 10-2, qui mesure la sensibilité sur 68 points dans un rayon de 10° autour du point de fixation. Cet examen, qui n'a donc pas été réalisé chez les patients du groupe nAC, aurait toutefois pu révéler une baisse de sensibilité plus subtile dans le CV central de ces patients, ce qui peut constituer une limite méthodologique.

Si le système prédictif est déficitaire chez les patients atteints de glaucome, cela suppose une désorganisation fonctionnelle des régions cérébrales impliquées dans la reconnaissance visuelle. Nous avons voulu amorcer l'étude de cette question dans l'Expérience 7, en étudiant de manière exploratoire l'activité corticale de patients atteints de glaucome et de participants contrôle lors de la catégorisation, en vision centrale, de stimuli visuels faiblement contrastés. Pour les raisons précédemment évoquées, nous n'avons conservé que la tâche de catégorisation de scènes d'intérieur *vs* d'extérieur.

3. Expérience 7 : Étude exploratoire en IRMf d'une désorganisation fonctionnelle cérébrale consécutive à une atteinte rétinienne glaucomateuse

3.1. Objectifs et hypothèses opérationnelles

L'Expérience 7 était une étude exploratoire en IRMf visant à étudier comment une perturbation du système prédictif pouvait se manifester au niveau cérébral. La littérature apporte de plus en plus de preuves de modifications structurelles et fonctionnelles dans le cerveau des patients atteints de glaucome. Des études montrent par exemple que la destruction progressive des cellules ganglionnaires de la rétine dans le glaucome peut déclencher une dégénérescence trans-synaptique dans le noyau géniculé latéral et le cortex visuel (Gupta & Yücel 2007; Weber, Chen, Hubbard, & Kaufman 2000; Yucel, Zhang, Weinreb, Kaufman, & Gupta 2001). Des preuves de plus en plus nombreuses provenant d'études en IRM appuient l'hypothèse d'importants changements anatomiques et fonctionnels du cortex liés à une dégénérescence neuronale dans le glaucome Boucard et al. (2009); Chen et al. (2013); Duncan, Sample, Weinreb, Bowd, et Zangwill (2007); Frezzotti et al. (2014); Gerente et al. (2015); Li et al. (2012); Qing, Zhang, Wang, et Wang (2010). Par exemple, une étude IRM de morphométrie (Boucard et al. 2009) révèle que la densité de la matière grise des patients atteints de GPAO est réduite dans la partie médiane du cortex occipital antérieur par rapport aux contrôles. Cela correspond approximativement aux projections des scotomes périphériques présents dans le champ visuel des patients. Dans une étude IRMf, Duncan et al. (2007) ont réalisé des cartographies rétinotopiques afin d'explorer la représentation du champ visuel dans le cortex visuel primaire des patients atteints de GPAO. Leurs résultats montrent des anomalies cérébrales fonctionnelles dans les régions du cortex visuel qui correspondent de façon rétinotopique à la perte de champ visuel périphérique. Les anomalies s'étendraient au delà du cortex visuel primaire. Plusieurs études ont mis en évidence des anomalies dans des régions de plus haut niveau du cortex visuel mais aussi dans des régions non visuelles (H. D. Brown, Woodall, Kitching, Baseler, & Morland 2016; Chen et al. 2013; Dacey 2004; Frezzotti et al. 2014; Li et al. 2012; Nelson et al. 1999). Par exemple, dans une étude en IRM, Chen et al. (2013) ont montré que la densité de la matière grise était réduite dans la scissure calcarine et le gyrus lingual, mais qu'elle était également diminuée ou augmentée dans de nombreuses régions corticales temporales, frontales et pariétales. De même, Frezzotti et al. (2014) ont montré une atrophie de la matière grise dans les régions visuelles de haut niveau de traitement comme le gyrus fusiforme latéral (au niveau du LOC). Cette même étude a aussi révélé une diminution de la connectivité

fonctionnelle dans le réseau cortical visuel, mais aussi dans le réseau de la mémoire de travail et dans le réseau de l'attention. Concernant spécifiquement le réseau cortical visuel, Dai et al. (2013) ont rapporté une altération de la connectivité fonctionnelle entre les aires visuelles précoces et les régions visuelles de plus haut niveau comme le gyrus fusiforme et le gyrus temporal inférieur. Dans l'ensemble, ces travaux mettent en évidence un large éventail de modifications structurelles et fonctionnelles dans le cerveau des patients glaucomateux. Néanmoins, à notre connaissance, aucune étude ne s'est intéressée aux activités cérébrales des patients en lien avec une tâche de reconnaissance visuelle. Il semble maintenant important d'étudier directement le fonctionnement de ces régions dans différentes situations requérant la reconnaissance de stimuli visuels complexes, comme par exemple avec notre paradigme de catégorisation de scènes. Une meilleure compréhension des changements fonctionnels cérébraux liés au glaucome apporterait non seulement des informations importantes pour les modèles de reconnaissance visuelle, mais permettrait aussi de décrire plus précisément les difficultés visuelles éprouvées par les patients, qui ne peuvent se résumer à de simples scotomes dans le champ visuel.

Dans l'Expérience 7, nous avons donc adapté le protocole de l'Expérience 6 à l'IRMf. Trois patients et quatre participants contrôles appariés en âge ont réalisé la tâche de catégorisation de scènes utilisée dans l'Expérience 6. En IRM, les stimuli étaient transmis sur un écran translucide situé derrière le scanner grâce à un vidéoprojecteur. Ce dispositif ne facilitant pas la perception de stimuli de très bas contraste, nous avons choisi de rehausser les niveaux de contraste des scènes par rapport à l'Expérience 6. Les valeurs de contraste RMS de 0.03 et 0.11 ont ainsi été utilisées. Notre objectif était d'explorer les réseaux cérébraux impliqués dans la reconnaissance de scènes visuelles chez quelques patients glaucomateux, et de les comparer à celui de participants contrôles. Nous étions particulièrement intéressées par les régions cérébrales qui pourraient être impliquées dans le modèle FP-péri de la reconnaissance visuelle. Ces régions sont **(i)** le cortex occipital postérieur dédié aux traitements visuels, **(ii)** les régions ventrales allant du cortex occipital à la partie inférieure du cortex temporal, dédiées aux traitements visuels de plus haut niveau, et **(iii)** la partie inférieure du cortex frontal (gyrus frontal inférieur, et notamment sa partie orbitofrontale), ainsi que le cortex insulaire, en lien avec les mécanismes prédictifs.

3.2. Méthode

3.2.1. Participants

Trois patients atteints de GPAO ont participé volontairement à cette étude. Quatre volontaires sains appariés en âge ont également participé à l'étude. Les participants contrôles et les patients faisaient partie de la cohorte de l'Expérience 6, et environ six mois séparaient les deux expériences. Les données démographiques, d'acuité visuelle et de périmétrie statique automatisée des trois patients et des quatre participants contrôles sont répertoriées dans le Tableau 3.4, et les champs visuels des patients sont visibles dans

la Figure 3.9. Les participants étaient testés sur un seul oeil. Une correction de la réfraction adaptée à l'IRM était utilisée si nécessaire pour que l'acuité soit maximale. Les participants ne présentaient pas d'autres pathologies oculaires, leur cristallin était clair et ils ne présentaient pas de pathologie neurologiques ou psychiatriques. Tous les participants ont donné leur consentement libre et éclairé avant de participer à l'expérience, qui a été menée en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine, et approuvée par un comité d'éthique (Comité de protection des personnes, Sud-Est V, ID/RCB : 2016-A00637-44).

Tableau 3.4: Caractéristiques et données cliniques des participants.
OG : oeil gauche ; OD : oeil droit ; F : femme ; H : homme. N/A : non applicable.

participant	âge	genre	oeil testé	acuité visuelle Monoyer	acuité visuelle Parinaud	indice de <i>Mean Deviation</i>	atteinte du champ visuel central	score MMSE
Glaucome 1	58	H	OG	10/10	2	- 4.11 dB	oui	26
Glaucome 2	59	H	OG	10/10	2	- 14.46 dB	oui	29
Glaucome 3	61	F	OG	10/10	2	- 4.67 dB	non	30
Contrôle 1	67	H	OG	10/10	2	N/A	N/A	30
Contrôle 2	62	H	OG	10/10	2	N/A	N/A	29
Contrôle 3	65	F	OD	10/10	2	N/A	N/A	29
Contrôle 4	64	F	OD	10/10	2	N/A	N/A	26

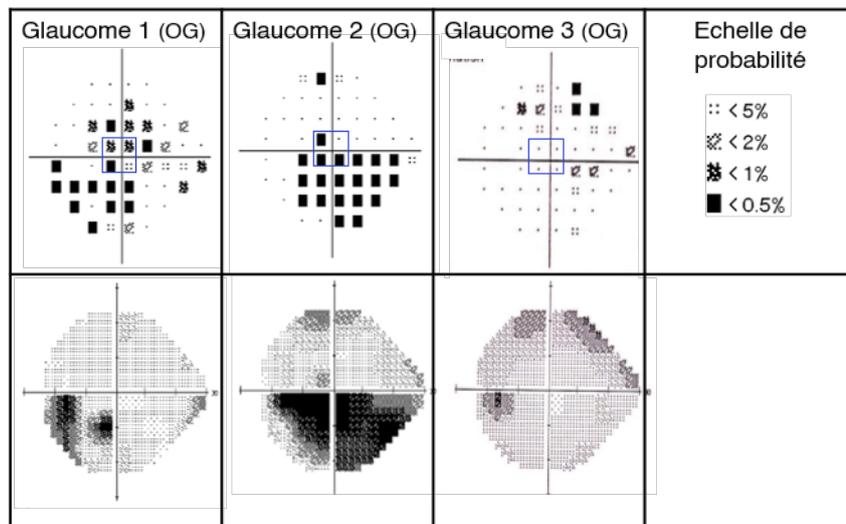


Figure 3.9: Champs visuels des patients testés dans l'Expérience 7. La première ligne montre les *probability plots* des déviations individuelles monoculaires des trois patients GPAO ayant réalisé l'étude d'IRMf. La seconde ligne montre la représentation en niveaux de gris de la sensibilité brute de chaque point testé (une région plus noire représente une sensibilité en dB plus faible). Ces informations périmétriques ont été obtenues avec une périmétrie statique automatisée de Humphrey (procédure SITA-Standard 24-2). L'oeil testé était l'oeil gauche pour les trois patients.

3.2.2. Stimuli et procédure

Les stimuli étaient des photographies de scènes en niveaux de gris issues de la même base ReCor que celles de la session SCENE de l’Expérience 6, et ces scènes avaient globalement les mêmes caractéristiques et ont été manipulées de la même manière que dans l’Expérience 6 (cf. sous-partie 2.1.2). Nous avons sélectionné 30 scènes (15 scènes d’intérieur et 15 scènes d’extérieur). Nous avons créé deux versions de chaque image en manipulant le contraste RMS : une version avec un RMS de 0.03 et une version avec un RMS de 0.11 (Figure 2.3). La luminance moyenne de toutes les images, ainsi que la luminance du fond gris uniforme sur lequel les stimuli étaient affichés, était de 72 pour une échelle de 256 niveaux de gris (soit 0.28 pour des valeurs de luminance comprises entre 0 et 1).



Figure 3.10: Exemples de scènes utilisées dans l’Expérience 7. Les images étaient présentées à deux niveaux de contrastes (RMS de 0.03 et 0.11).

L’expérience était programmée avec E-Prime 2. La taille angulaire des scènes était d’environ 6° d’angle visuel de largeur. Comme dans l’Expérience 6, les participants réalisaient une tâche de catégorisation : ils devaient appuyer sur un dispositif de réponse lorsque l’image appartenait à la catégorie cible, et retenir leur réponse lorsqu’elle appartenait à la catégorie distractive, dans un paradigme go/no-go. Ils donnaient leur réponse à l’aide d’un boîtier de réponse placé à l’intérieur du scanner. À chaque essai, l’exactitude de la réponse et le temps de réponse étaient enregistrés.

L’expérience comprenait deux scans fonctionnels conçus avec un paradigme IRM de type bloc. Un scan fonctionnel était composé de six blocs de tâche (trois blocs pendant lesquels le contraste RMS des scènes était de 0.03 et trois blocs pendant lesquels le contraste RMS était de 0.11), et de six blocs de repos pendant lesquels un point de fixation central était présenté sur le fond gris uniforme. Chaque bloc de tâche durait 25 secondes et permettait l’acquisition de 10 volumes fonctionnels (un toutes les 2.5 secondes) correspondant à 10 essais expérimentaux. Chaque bloc de repos durait 10 secondes et permettait l’acquisition de quatre volumes fonctionnels (un toutes les 2.5

secondes). Ainsi, pour chaque scan fonctionnel, 84 volumes fonctionnels étaient acquis (durée d'acquisition des scans fonctionnels : environ 10 minutes). Au total, il y avait 60 essais expérimentaux par scan fonctionnel, soit 120 au total (60 essais go et 60 essais no-go).

3.2.3. Acquisition et traitement des données IRMf

Les acquisitions ont été réalisées sur un scanner Philips Achieva 3T (Philips Medical Systems, Eugene, OR). Avant le premier essai de chaque session d'acquisition, cinq mesures du volume (*dummies scans*) étaient réalisées pour permettre la stabilisation du champ magnétique. Les acquisitions fonctionnelles (images pondérées en T2*) étaient ensuite réalisées en mode EPI afin d'exploiter le contraste BOLD. Quarante-deux coupes adjacentes de 3 mm d'épaisseur, parallèles au plan bi-commissural, étaient acquises de manière séquentielle de bas en haut en incluant le cervelet (TR = 2500 ms, TE = 30 ms, angle de basculement = 82°, champ de vue = 240 × 240 × 126 mm, matrice d'acquisition = 80 × 80, matrice de reconstruction = 80 × 80, taille des voxels = 3 × 3 × 3 mm). Après les scans fonctionnels, un scan anatomique (images pondérées en T1, champ de vue = 256 × 224 × 175 mm, résolution = 1.33 × 1.75 × 1.37 mm, matrice d'acquisition = 192 × 115 × 128, matrice de reconstruction = 288 × 288 × 128) était acquis afin de pouvoir superposer les images fonctionnelles à une image anatomique de haute résolution spatiale et de pouvoir identifier les régions cérébrales par rapport à des atlas anatomiques de référence.

Les données d'IRMf ont été traitées en utilisant le programme *Statistical Parametric Mapping 12* (SPM, Wellcome Trust Centre for Neuroimaging, Londres, R.U.) dans MATLAB. Le traitement spatial des volumes fonctionnels de chaque participant a été réalisé selon les étapes suivantes. Nous avons tout d'abord effectué un *slice-timing* afin de corriger le décalage temporel d'acquisition entre les coupes composant un même volume fonctionnel. Nous avons ensuite corrigé les artefacts dus aux mouvements de la tête du participant (rotations et translations) en réalignant chaque volume fonctionnel au premier volume du scan. À l'issue du réalignement, une image moyenne des images fonctionnelles a été créée. Nous avons alors appliqué une coregistration qui permet de faire correspondre les orientations géométriques de l'image anatomique à la moyenne des images fonctionnelles. L'image anatomique a été ensuite normalisée dans l'espace MNI à l'aide d'un *template* anatomique, puis, les images fonctionnelles ont été normalisées à l'aide des paramètres calculés lors de la normalisation de l'image anatomique. Enfin, les images fonctionnelles ont été lissées par un filtre gaussien afin d'augmenter le rapport signal sur bruit.

Les données fonctionnelles ont ensuite été modélisées à l'aide d'un modèle linéaire général (Friston et al. 1995). Pour chaque participant, nous avons créé une matrice de dessin incluant trois régresseurs correspondant aux différentes conditions expérimentales (Catégorisation/RMS-0.03, Catégorisation/RMS-0.11, Repos). Pour chaque régresseur, les périodes correspondant à chacun des blocs fonctionnels ont chacune été convoluées

avec à une fonction de réponse hémodynamique canonique de type *box-car*. Les paramètres de mouvement (trois rotations et trois translations) ont également été modélisés dans la matrice de dessin comme régresseurs de non intérêt. Nous avons ensuite généré des cartes d'activation pour chaque participant à partir des contrastes entre les valeurs des paramètres estimés pour les différentes conditions expérimentales. Plus spécifiquement, nous avons généré trois cartes d'activation pour chaque participant correspondant aux contrastes statistiques suivants : [Catégorisation > Repos], [Catégorisation-RMS0.03 > Repos], et [Catégorisation-RMS0.11 > Repos]. Puisque nous étions spécifiquement intéressés par le réseau cérébral supposément impliqué dans le modèle fréquentiel prédictif, nous avons restreint l'analyse aux seules régions concernées. Dans cet objectif, nous avons défini des régions d'analyse a priori en créant un masque inclusif grâce à la boîte à outils WFU_PickAtlas sur MATLAB (https://www.nitrc.org/projects/wfu_pickatlas). Ce masque permettait d'analyser les cartes de contraste dans nos régions d'intérêt à savoir le lobe occipital, les gyri fusiformes et parahippocampiques, les gyri temporaux inférieurs, les gyri frontaux inférieurs (incluant leurs parties *triangulaires*, *opercularis* et *orbitaris*) et l'insula. Nous avons retenu les clusters d'activation sur la base de l'intensité de la réponse individuelle des voxels (à un seuil de 0.001) et de l'étendue spatiale de l'activation (clusters d'au moins 10 voxels)

3.3. Résultats

3.3.1. Performances comportementales

Les performances (d') de chaque participant sont illustrées dans la Figure 3.11 ci-dessous. On remarque que les participants contrôles ont globalement une très bonne sensibilité pour catégoriser les scènes (d' moyens = 4.68, 4.79, 4.72, 3.66). Pour trois d'entre eux, Contrôle 1, Contrôle 2 et Contrôle 3, les performances sont bonnes quel que soit le niveau de contraste. La participante Contrôle 4 montre une légère baisse de sensibilité pour le plus haut niveau de contraste (RMS = 0.11 ; $d' = 2.94$), mais ses performances sont très bonnes pour le plus bas niveau (RMS = 0.03 : $d' = 4.37$), ce qui laisse penser que la mesure a pu être peu précise chez cette participante. Les patients ont quant à eux des résultats plus variés. Le patient Glaucome 2 a une sensibilité faible (d' autour de 1) quel que soit le niveau de contraste, témoignant d'une difficulté à catégoriser les scènes chez ce patient dont la vision centrale est particulièrement atteinte. Le patient Glaucome 1 a une bonne sensibilité lorsque le contraste RMS est de 0.11 ($d' = 4.24$) mais ses performances diminuent drastiquement lorsque le contraste RMS est de 0.03 ($d' = 1.27$). La patiente Glaucome 3 a globalement une bonne sensibilité, qui équivaut à celle de Contrôle 4 ($d' = 3.66$). Ses performances sont meilleures lorsque le contraste RMS est de 0.03 ($d' = 4.51$) que lorsqu'il est de 0.11 ($d' = 2.8$).

On retiendra surtout qu'au moins deux des trois patients, Glaucome 1 et Glaucome 2, présentent des difficultés pour catégoriser les scènes. La conclusion est moins évidente chez Glaucome 3. La tâche ne semble au contraire pas poser de difficulté aux participants contrôles.

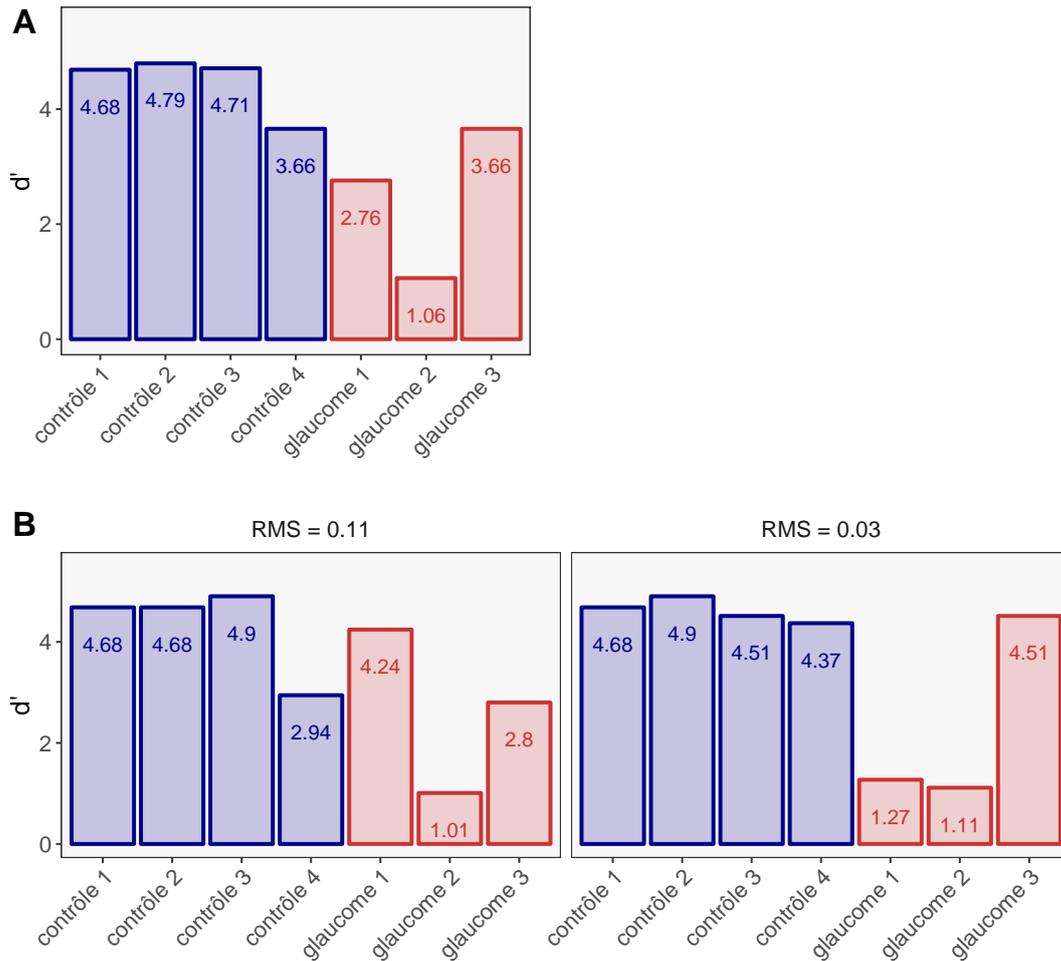


Figure 3.11: Résultats comportementaux de l'Expérience 7 : indice d' pour chaque participant **A.** moyenné sur les deux niveaux de contraste et **B.** pour chaque niveau de contraste (RMS-0.11 ; RMS-0.03).

3.3.2. Analyse des données IRMf

L'objectif principal de ces analyses était d'identifier les régions cérébrales impliquées dans la reconnaissance visuelle des scènes en utilisant le contraste [Catégorisation > Repos] en considérant d'abord les deux niveaux de contraste RMS, puis en considérant séparément les conditions RMS-0.03 et RMS-0.11. Pour chaque participant, les coordonnées du pic d'activation des clusters sont en Annexe D. Les résultats sont résumés ci-dessous en prenant en compte les clusters d'activation significatifs dans les régions **(i)** occipitales, **(ii)** occipito-temporales et **(iii)** frontales.

Dans un premier temps, nous avons examiné le contraste [Catégorisation > Repos] en considérant les deux niveaux de contraste RMS (Figure 3.12). Ce contraste révèle des activations communes aux patients glaucomateux et aux participants contrôles, surtout dans le cortex occipital. Par exemple, tous activent bilatéralement les régions occipitales

postérieures et le gyrus occipital moyen. Dans le cortex occipito-temporal, nous pouvons observer que trois participants contrôles sur les quatre testés activent fortement et bilatéralement la PPA, au niveau du gyrus parahippocampique. En comparaison, seul un des trois patients testés (Glaucome 1) active cette région, et dans l'hémisphère droit uniquement. Au niveau du lobe frontal, on constate d'abord que la partie *triangularis* du gyrus frontal inférieur est impliquée chez tous les participants, sauf chez le patient Glaucome 2. Ce patient présente par ailleurs de mauvaises performances comportementales et on peut noter qu'il n'a aucune activation dans le lobe frontal. De manière intéressante, les participants contrôles activent systématiquement le cortex orbitofrontal (partie *orbitalis* du gyrus frontal inférieur) - ou, pour Contrôle 1, le cortex insulaire, adjacente à la région orbitofrontale. Parmi les patients, Glaucome 1 et Glaucome 2 n'activent pas ces régions. Glaucome 3 active en revanche fortement et bilatéralement l'insula et le cortex orbitofrontal. Il est intéressant de noter que cette patiente est par ailleurs la seule qui maintient de bonnes performances comportementales.

Nous avons ensuite décomposé le contraste [Catégorisation > Repos] en l'examinant pour chaque niveau de contraste (Figure 3.13). Là encore, à la fois les participants contrôles et les patients présentent des activations au niveau des régions occipitales postérieures, et ce pour les deux niveaux de contraste, à l'exception du patient Glaucome 2 pour lequel ces régions s'activent seulement pour le niveau de contraste le plus fort (RMS-0.11). Ce patient a une forte atteinte du champ visuel central et ses performances comportementales témoignent par ailleurs de grosses difficultés à catégoriser les scènes.

On retrouve les activations orbitofrontales-insulaires chez les quatre participants contrôles. Ces activations sont observées, à la fois pour le contraste RMS 0.03 et RMS 0.11 pour les Contrôles 1, 2 et 4, mais seulement pour le contraste RMS 0.03 (contraste bas) pour le Contrôle 3. Cette analyse nous permet aussi de constater une légère activation dans ces régions orbitofrontales/insulaires pour le patient Glaucome 1 lorsque le contraste RMS est de 0.03. Chez le patient Glaucome 3, qui a de bonnes performances comportementales, ces régions sont activées pour les deux niveaux de contraste, mais tout particulièrement lorsque le contraste RMS est de 0.03.

Au niveau du cortex parahippocampique, les résultats sont par contre trop divergents pour en tirer des conclusions. Les quatre contrôles montrent une activation de la PPA au moins pour un niveau de contraste. Cette région est également activée chez les patients Glaucome 1 et Glaucome 2, et ce surtout lorsque le contraste est relativement fort (RMS 0.11).

3.4. Discussion

Bien que cette étude IRMf apporte des résultats très hétérogènes, elle apporte toutefois des pistes intéressantes pour orienter les hypothèses de futures études. Nous l'avons évoqué dans les différents chapitres, l'implémentation neurale du modèle FP/FP-peri serait distribuée dans différentes régions corticales. Le système de reconnaissance visuelle serait implémenté dans les cortex occipital et occipito-temporal ventral. Le système de

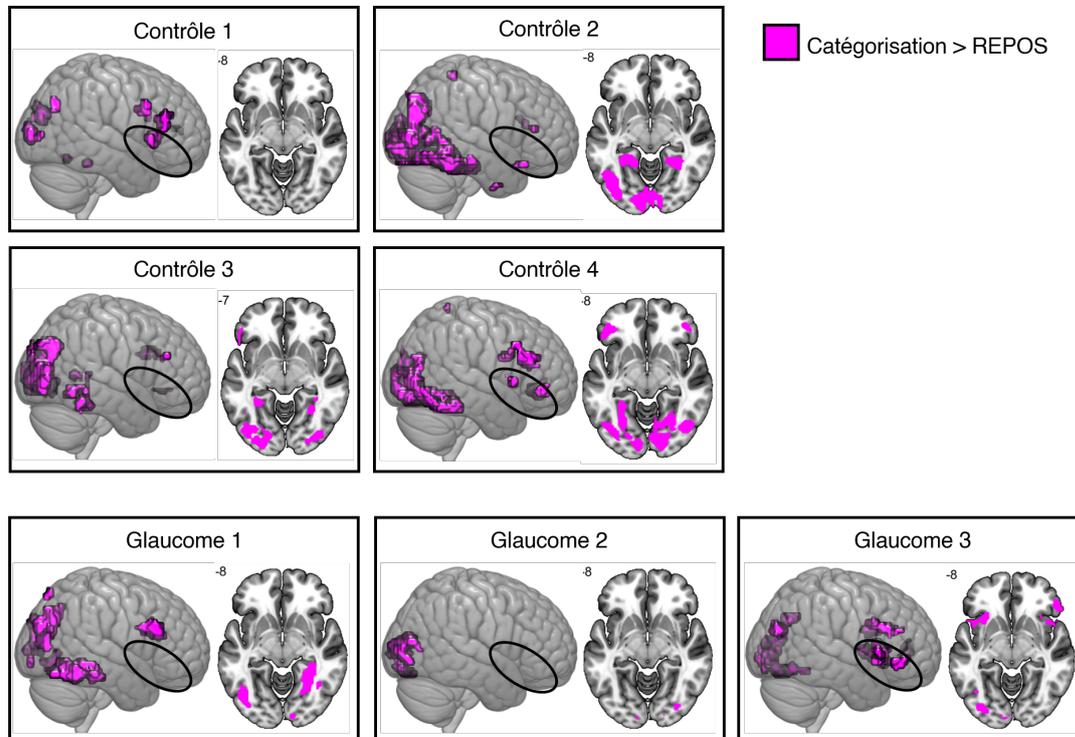


Figure 3.12: Activations fonctionnelles des participants contrôles et des patients atteints de glaucome lors de la tâche de catégorisation de scènes (contraste [Catégorisation > Repos]). La cerveau 3D permet d'avoir une vue d'ensemble sur le réseau mis en jeu dans la catégorisation des scènes. Les ellipses noires indiquent approximativement l'emplacement du cortex orbitofrontal et du cortex insulaire. On constate une activation dans ces régions chez chaque participant contrôle (N.B. le cluster est moins visible pour le Contrôle 3 car situé seulement dans l'hémisphère gauche qui apparaît ici en transparence). On ne retrouve pas d'activation dans ces régions frontales pour Glaucome 1 et Glaucome 2, et elles sont au contraire très activées chez Glaucome 3, la seule patiente qui maintient de bonnes performances comportementales. Les coupes axiales ($z = -8$) permettent de révéler les activations de la PPA, dans le cortex parahippocampique. Trois participants contrôles sur les quatre activent bilatéralement la PPA. Seul le patient Glaucome 1 active cette région, et dans l'hémisphère droit uniquement.

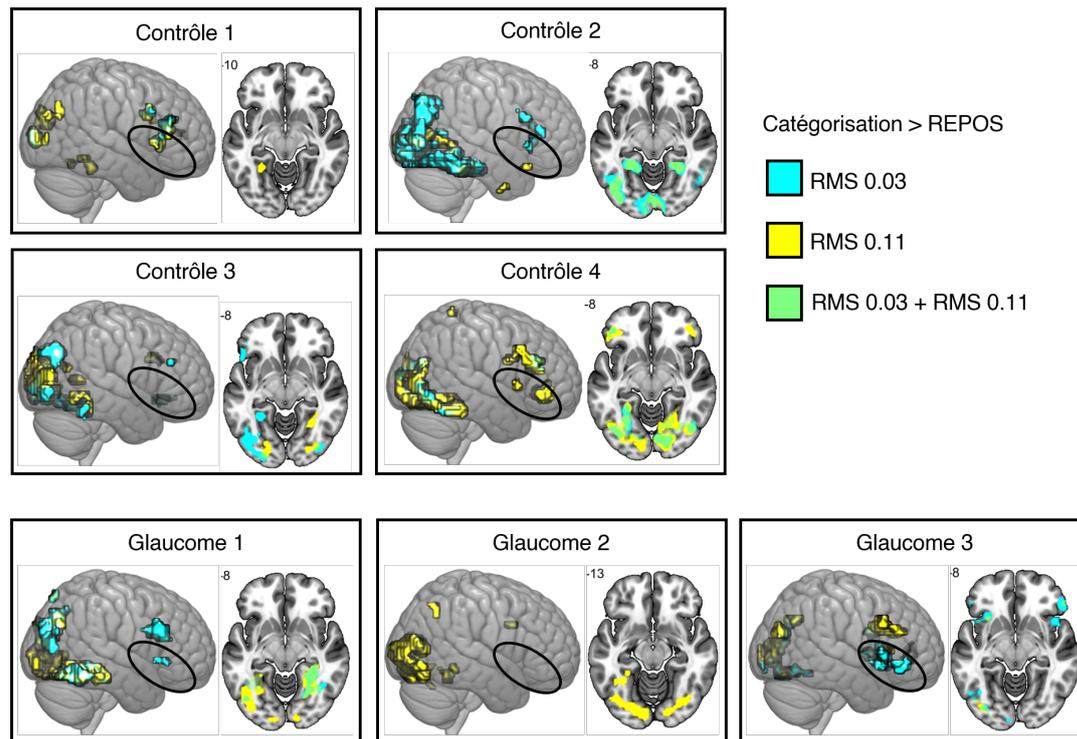


Figure 3.13: Activations fonctionnelles des participants contrôles et des patients atteints de glaucome lors de la tâche de catégorisation de scènes (contraste [Catégorisation > Repos]), en fonction du niveau de contraste des scènes. Le cerveau 3D permet d'avoir une vue d'ensemble sur le réseau mis en jeu dans la catégorisation des scènes. Les ellipses noires indiquent approximativement l'emplacement du cortex orbitofrontal et du cortex insulaire. Les coupes axiales ($z = -8$ environ) permettent de révéler les activations de la PPA, dans le cortex parahippocampique.

gestion des prédictions pourrait être implémenté dans le cortex frontal inférieur, et tout particulièrement dans sa partie orbitofrontale. Nous avons fait l'hypothèse que l'altération irréversible de la vision périphérique dans le glaucome perturberait le système de reconnaissance dans son ensemble et que cela se traduirait par des modifications fonctionnelles au niveau cérébral. Ici, nous avons réalisé des scans fonctionnels de trois patients glaucomateux et quatre participants contrôles tandis qu'ils catégorisaient des scènes présentées en vision centrale. Les résultats indépendamment du niveau de contraste pointent vers deux phénomènes possibles. D'abord, la PPA, cette région fonctionnellement définie comme étant spécifiquement impliquée dans le traitement des scènes visuelles (plutôt que d'autres stimuli plus simples) semble moins recrutée chez les patients que chez les participants contrôles. Seul un patient active la PPA, et ce seulement dans l'hémisphère droit. Le glaucome pourrait ainsi perturber la manière dont la PPA est impliquée lors de la perception de scènes. On rappelle d'ailleurs que cette région est plus sensible à des stimulations dans le champ visuel périphérique que central (Arcaro et al. 2009; Baldassano et al. 2016; Hasson et al. 2002; Levy et al. 2001; Malach et al. 2002). Avec le glaucome, le cortex reçoit avec le temps de moins en moins d'information provenant de la rétine périphérique, ce qui pourrait mettre cette région dans un 'sous-régime' de fonctionnement. Un dysfonctionnement dans la PPA pourrait alors en partie expliquer les difficultés de reconnaissance de scènes chez les patients glaucomateux, même quand celles-ci apparaissent dans leur vision centrale résiduelle.

Ensuite, nous avons observé une activation du cortex orbitofrontal et/ou du cortex insulaire antérieur (adjacente au cortex orbitofrontal) chez tous les participants contrôles. Ce n'est pas le cas chez deux des patients glaucomateux testés (Glaucome 1 et Glaucome 2). La troisième patiente en revanche (Glaucome 3) active fortement et bilatéralement cette région. À cet égard, il est intéressant de remarquer que cette patiente montre de bonnes performances comportementales (d' moyen = 3.66) qui sont comparables à celles des participants contrôles (notamment Contrôle 4, d' moyen = 3.66). Dans le cadre d'un modèle FP-peri, nous avons fait l'hypothèse que le glaucome perturberait le système de reconnaissance visuelle dans son ensemble. Notamment, le système de prédictions (Figure 1.21) de moins en moins alimenté par un signal BF's provenant de la rétine périphérique, pourrait être moins efficace. Les observations faites ici sont plutôt compatibles avec cette hypothèse. Si le cortex orbitofrontal est impliqué dans la gestion des prédictions visuelles, il est alors possible que cette région dysfonctionne (comme pour les patients Glaucome 1 et Glaucome 2), ou qu'elle soit surinvestie dans une logique de stratégie compensatoire amenant à de bonnes performances de catégorisation (comme pour la patiente Glaucome 3). Ainsi, les résultats de l'Expérience 7 ouvrent des pistes intéressantes pour étudier au niveau neurofonctionnel l'impact du glaucome sur les mécanismes de reconnaissance visuelle. Les patterns observés en tant que tels sont à prendre avec précaution car ils ne concernent que très peu de participants. Par ailleurs, on constate que lorsque l'on regarde les résultats en fonction du niveau de contraste, ces patterns ne sont plus aussi évidents, surtout au niveau de la PPA, ce qui est possiblement lié à un nombre d'essais trop faible lorsqu'on ne considère plus qu'un niveau de contraste. Les observations faites

3. Expérience 7 : Étude exploratoire en IRMf d'une désorganisation fonctionnelle cérébrale consécutive à une atteinte rétinienne glaucomateuse

dans l'Expérience 7 sont alors surtout à considérer comme permettant de générer de nouvelles hypothèses. En l'améliorant, ce protocole simple de catégorisation de scènes pourrait suffire à mettre en évidence des différences de fonctionnement dans des régions du modèle FP, comme la PPA ou le cortex orbitofrontal.

Chapitre 4

Discussion générale

Dans le Chapitre 1 de ce manuscrit de thèse, nous avons formalisé un modèle théorique de la reconnaissance visuelle : le modèle fréquentiel prédictif (FP). Ce modèle permet de résumer et concilier les différentes propositions théoriques dans le domaine de la perception visuelle de stimuli complexes. Ce modèle repose d'abord sur le postulat de base d'une analyse *coarse-to-fine* de la scène visuelle. L'analyse ultra-rapide des BFs permettrait l'émergence d'une représentation rudimentaire de la scène visuelle que l'on peut appeler le *gist* de la scène. Le *gist* permettrait d'avoir accès à la configuration spatiale de la scène et à des percepts visuels d'ensemble (sous forme de *summary statistics*), et permettrait ainsi une première catégorisation du stimulus à un niveau sur-ordonné. Cette première représentation serait ensuite enrichie avec l'accumulation progressive de l'information précise et détaillée contenue dans les HFs.

Le second point d'ancrage théorique du modèle FP se situe dans le domaine du codage prédictif. L'émergence ultra-rapide du *gist* permettrait d'activer des mécanismes prédictifs. Des prédictions seraient alors rapidement formées et renvoyées sur le système de reconnaissance visuelle afin de faciliter l'intégration ascendante des HFs et leur interprétation. Comme nous l'avons vu, à la fois l'hypothèse d'un principe *coarse-to-fine* et l'hypothèse de l'existence de mécanismes prédictifs sont appuyées par de nombreuses données empiriques (cf. Chapitre 1, partie 2). Par ailleurs, le modèle FP permet de rendre compte de l'extrême rapidité (traitement ultra-rapide du *gist*), mais aussi de l'efficacité et de la fiabilité de la reconnaissance visuelle. Nous avons ensuite argumenté que l'émergence du *gist* de la scène pourrait être particulièrement liée à l'analyse de l'information provenant du champ visuel périphérique (Chapitre 1, partie 3). En effet, la majorité du signal en BFs est extraite en vision périphérique. Par ailleurs, l'information en vision périphérique pourrait être rapidement traitée afin de former des prédictions qui guideraient ensuite le traitement des HFs en vision centrale. Malheureusement, peu d'études permettent d'étayer cette hypothèse, la majorité utilisant des petits stimuli présentés en vision centrale.

L'objectif de cette thèse était alors de tester l'hypothèse d'un modèle FP-péri, c'est-à-dire de tester comment l'information en vision périphérique peut influencer la reconnaissance visuelle en vision centrale dans le contexte du modèle FP. Dans le Chapitre 2, nous avons présenté une série d'expériences qui nous a permis de tester cette hypothèse chez le sujet sain avec un paradigme d'interférence. Dans le Chapitre 3, nous avons testé cette hypothèse en considérant le glaucome comme modèle pathologique de la vision périphérique. Dans ce dernier Chapitre, après avoir fait une rapide synthèse des résultats, nous présenterons en quoi les résultats de ce travail de thèse peuvent étayer et préciser le modèle FP-péri, puis nous proposerons des perspectives de recherche.

1. Apports des résultats à la théorie

1.1. Synthèse des résultats obtenus

Expérience 1 : Cette expérience a tout d'abord confirmé, dans la logique du principe *coarse-to-fine*, que les scènes filtrées en BFs étaient catégorisées plus rapidement et avec plus de précision que les scènes filtrées en HFes à la fois lorsqu'elles étaient présentées seules et lorsqu'elles étaient dans des stimuli hybrides. Ensuite, les résultats ont montré que bien qu'il faille un temps important pour reconnaître l'information en HFes dans les scènes hybrides, celle-ci influence considérablement la catégorisation de l'information en BFs. Nous avons en effet observé un effet d'interférence des HFes sur la catégorisation des BFs. Contrairement à nos hypothèses, l'interférence des HFes était plus forte que l'interférence des BFs, et ce quel que soit le temps de présentation du stimulus hybride (30 ms ou 100 ms). Nous avons interprété ce résultat comme étant la résultante de contraintes liées à la tâche. La catégorisation de scènes hybrides - un stimulus particulièrement ambigu - pourrait nécessiter une stratégie d'analyse fine qui mettrait plus de poids sur l'analyse des HFes. Mais les mécanismes d'une telle stratégie restent flous et, de notre point de vue, cet important et inattendu effet d'interférence des HFes sur la catégorisation des BFs souligne surtout le caractère non écologique des scènes hybrides. Dans cette expérience, nous avons également mesuré comment la similarité physique entre les scènes HFes et BFs des hybrides (spectres d'amplitude et configurations spatiales semblables) pouvait moduler l'effet d'interférence. Cette question avait surtout son importance pour étudier la nature de l'information en BFs (sémantique et/ou physique) qui interfère sur la catégorisation des HFes. Dans cette condition, nous n'avons trouvé aucune preuve d'une modulation de l'effet d'interférence par la similarité physique. Ces résultats ne permettent alors pas d'appuyer l'effet de la similarité physique de [Mu et Li \(2013\)](#), mais restent toutefois cohérent avec l'étude de [Kauffmann et al. \(2015c\)](#), qui avec la même base de stimuli, n'avaient pas non plus observé d'effet de la similarité physique. Mise à part le fait que notre équipe de recherche n'utilise pas les mêmes stimuli que [Mu et Li \(2013\)](#), les facteurs à l'origine de cette divergence de résultats restent à clarifier. On peut remarquer, par exemple, que le nombre d'erreurs de catégorisation est très important dans notre expérience. Il est donc possible que la difficulté de la tâche vienne masquer les effets en lien avec la similarité physique. Cette interprétation est d'autant plus plausible que nous avons observé un effet de la similarité physique dans les Expériences 2 et 3 de cette thèse.

Expérience 2 : Cette expérience nous a permis de montrer que la catégorisation d'une scène en vision centrale était influencée par la présence d'une scène en vision périphérique. Ce résultat suggère que l'information périphérique est traitée automatiquement et que cette information est intégrée au processus de reconnaissance de l'information en vision centrale. Par ailleurs, nous n'avons observé cet effet d'interférence périphérique que dans le cas où les deux scènes partageaient des propriétés physiques similaires (spectres

d'amplitude et configurations spatiales semblables). On peut cette fois-ci remarquer que cette tâche était plus facile que celle utilisée dans l'Expérience 1 avec les scènes hybrides, favorisant possiblement la mise en évidence d'un effet de la similarité physique.

Expérience 3 : Dans cette expérience, nous avons adapté l'Expérience 2 à la technique de l'IRMF afin de pouvoir coupler nos mesures comportementales à des mesures neurobiologiques. Nous avons d'abord répliqué les résultats comportementaux de l'Expérience 2 sur les temps de réponse correctes. Les effets étaient par contre moins consistants entre les deux expériences pour la proportion de réponses correctes, ce qui est probablement dû à la taille de l'échantillon, plus petite dans l'Expérience 3 que dans l'Expérience 2. L'analyse des données IRMF a mis en évidence que l'effet d'interférence périphérique s'accompagnait d'une activation bilatérale du cortex orbitofrontal qui s'étendait jusqu'à l'insula antérieure. Autrement dit, cette région était plus activée lorsque les scènes centrale et périphérique étaient incongruentes que congruentes. Cette activation n'était observée que lorsque les deux scènes avaient des propriétés physiques similaires, soulignant de nouveau le rôle des propriétés physiques des stimuli dans l'effet d'interférence périphérique.

Expérience 4 : Cette expérience nous a permis de montrer que la catégorisation d'un objet en vision centrale était influencée par la présence de l'information contextuelle d'une scène présentée en vision périphérique. Ce résultat suggère que cette information contextuelle peut être intégrée aux processus de reconnaissance d'objets en vision centrale. Par contre, nous n'avons pas observé cet effet lorsque la scène et l'objet étaient perçus simultanément. Il fallait que la scène soit perçue légèrement en amont (30 ms ou 150 ms) pour qu'elle influence la reconnaissance de l'objet. Nos modèles statistiques ont également montré que l'influence de la scène contextuelle (i.e. l'effet d'interférence) augmentait linéairement avec le temps de pré-exposition de la scène. Comme nous le verrons dans la suite de ce chapitre, c'est ce résultat en particulier qui conforte notre hypothèse d'une influence de mécanismes prédictifs issus de la vision périphérique.

Expérience 5 : Avec cette expérience, nous avons mesuré l'effet d'interférence avec une nouvelle mesure. Nous avons diminué la visibilité de l'objet de manière paramétrique en manipulant un paramètre qui permet d'ajuster la cohérence de la phase de l'objet. Cette manipulation paramétrique nous a permis d'estimer des fonctions psychométriques liant la précision de la catégorisation au niveau de visibilité de l'objet. Nous avons montré que les seuils de visibilité étaient plus bas lorsque le contexte périphérique était congruent que lorsqu'il était incongruent. Ce résultat suggère qu'une information utile/prédictive en vision périphérique peut favoriser la reconnaissance d'objets peu visibles dans la scène. Par ailleurs, les résultats de cette expérience montrent qu'il n'y avait pas d'effet d'interférence lorsque la scène périphérique était perçue dans sa version *scrambled*, c'est-à-dire quand l'information de phase de l'image était brouillée et que seule l'information issue du spectre d'amplitude (fréquences spatiales et orientations) était disponible. Cette expérience nous amène à nous questionner dans la suite de ce chapitre sur la nature des caractéristiques physiques utilisées pour activer des prédictions (spectre d'amplitude *vs*

configuration spatiale).

Expérience 6 : Dans cette expérience, nous avons observé que des patients atteints de glaucome ayant une atteinte périmétrique au niveau du champ visuel périphérique présentent des déficits subtils pour catégoriser des scènes présentées à très faible contraste en vision centrale, pourtant considérée comme intacte chez ces patients. Ce résultat montre que la seule atteinte de la vision périphérique peut perturber les mécanismes de reconnaissance en vision centrale, confortant à nouveau le rôle de la vision périphérique dans le système prédictif.

Expérience 7 : Cette expérience est une étude exploratoire en IRMf. Son objectif était d'étudier le réseau cérébral impliqué dans la reconnaissance de scènes chez les patients glaucomateux, et plus particulièrement la désorganisation cérébrale induite par l'atteinte de la rétine périphérique. Les résultats préliminaires montrent des différences d'activation entre les patients et des participants sains dans des régions cruciales du modèle FP, comme le cortex orbitofrontal ou la PPA. Les résultats sont à prendre avec précaution car nous n'avons scanné que quatre participants. En revanche, ils permettent de faire l'hypothèse qu'une désorganisation du système prédictif chez les patients glaucomateux pourrait se manifester au niveau du cortex orbitofrontal ou insulaire, et du cortex inférotemporal.

En conclusion, les expériences du Chapitre 2 réalisées chez le sujet jeune et sain ont montré de manière consistante une influence de l'information périphérique sur la reconnaissance visuelle en vision centrale. Les scènes (Expériences 2 et 3) ou les objets (Expériences 4 et 5) perçus en vision centrale étaient mieux reconnus lorsque l'information périphérique était sémantiquement congruente que lorsqu'elle était incongruente. L'effet d'influence de la vision périphérique semble reproductible et assez robuste. Cependant, comme nous le discuterons dans la sous-partie 1.2 de ce chapitre, les effets sont de faible magnitude. Par ailleurs, les expériences du Chapitre 3 menées auprès de patients atteints de glaucome indiquent que la perte de la vision périphérique perturberait la reconnaissance visuelle dans son ensemble. Cette perturbation pourrait être étudiée au niveau cérébral à travers l'étude de la réorganisation fonctionnelle des régions cérébrales impliquées dans un modèle FP-péri de la reconnaissance visuelle.

1.2. Taille de l'effet d'interférence périphérique et importance pratique

Le Tableau 4.1 regroupe toutes les tailles d'effets de l'effet de congruence / d'interférence des différentes expériences du Chapitre 2 (i.e. OR et %change dérivés des β des modèles statistiques). Le tableau présente ensuite des exemples de différences prédites par les modèles statistiques pour des valeurs de temps de réponses et de proportion de

réponses correctes fixées arbitrairement pour la condition congruente¹. À titre d'illustration, pour l'effet d'interférence des BFs sur le traitement des HFs dans les scènes hybrides, le modèle statistique de l'Expérience 1 (Session 2, première ligne du Tableau 4.1) prédit que si - dans une future expérience - le temps de réponse moyen pour la condition congruente est de 580 ms, alors le temps de réponse moyen pour la condition incongruente devrait être de 602 ms.

Tableau 4.1: Taille de l'effet d'interférence dans les expériences du Chapitre 2. Le tableau récapitule les pourcentages de changement (*%change*, pour les temps de réponses correctes) et les *Odds Ratio* (OR, pour les proportions de réponses correctes) dérivés des β des modèles statistiques. Pour chaque effet, le temps de réponses correctes et/ou la proportion de réponses correctes est fixée arbitrairement pour la condition congruente et correspond à la prédiction du modèle pour la condition incongruente.

		<i>%change</i>	congruent (temps de réponse correcte moyen fixé à 580 ms)	incongruent (temps de réponse correcte moyen prédit par le modèle)	OR	congruent (probabilité de faire une réponse correcte fixée à 0.90)	incongruent (probabilité de faire une réponse correcte prédite par le modèle)
Interférence des BFs sur les HFs en VC	Expérience 1	3.78 %	580 ms	602 ms (+22 ms)	1.47	0.90	0.86
Interférence d'une scène en VP sur la catégorisation de scènes en VC - effet global	Expérience 2	2.29 %	580 ms	593 ms (+13 ms)	1.29	0.90	0.87
	Expérience 3	1.80 %	580 ms	590 ms (+10ms)	N/A	N/A	N/A
Interférence d'une scène en VP sur la catégorisation de scènes en VC - condition PHYSIM	Expérience 2	3.78 %	580 ms	602 ms (+22 ms)	N/A	N/A	N/A
	Expérience 3	3.22 %	580 ms	599 ms (+19 ms)	1.43	0.90	0.86
Interférence d'une scène en VP sur la catégorisation de scènes en VC - condition PHYDIS	Expérience 2	N/A	N/A	N/A	N/A	N/A	N/A
	Expérience 3	N/A	N/A	N/A	N/A	N/A	N/A
Interférence d'une scène en VP sur la catégorisation d'objets en VC - effet global	Expérience 4	0.87 %	580 ms	585 ms (+5 ms)	1.30	0.90	0.87
Interférence d'une scène en VP sur la catégorisation d'objets en VC - SOA = 30 ms	Expérience 4	0.61 %	580 ms	584 ms (+4 ms)	N/A	N/A	N/A
Interférence d'une scène en VP sur la catégorisation d'objets en VC - SOA = 150 ms	Expérience 4	1.80 %	580 ms	590 ms (+10 ms)	N/A	N/A	N/A
Interférence d'une scène en VP sur la catégorisation d'objets (visibilité réduite) en VC - SOA = 150 ms (condition scènes intactes)	Expérience 5	2.55 %	580 ms	594 ms (+14 ms)	N/A	N/A	N/A
Interférence d'une scène en VP sur la catégorisation d'objets (visibilité réduite) en VC - SOA = 150 ms (condition scènes <i>scrambled</i>)	Expérience 5	N/A	N/A	N/A	N/A	N/A	N/A

Avec le Tableau 4.1, on peut constater que la taille de l'effet d'interférence de la vision périphérique mesurée dans les Expériences 2 à 5 est comparable à celle de l'effet d'interférence des BFs sur la catégorisation des HFs en vision centrale mesurée dans l'Expérience 1 lorsque des informations physiquement similaires sont présentées dans

1. Les valeurs pour la condition congruente n'ont pas été fixées tout à fait arbitrairement : un temps de 580 ms et une proportion de réponses correctes de 0.90 correspondent à peu près au temps de réponses correctes moyen et à la proportion de réponses correctes moyenne à travers toutes les expériences du Chapitre 2.

les champs visuels central et périphérique (Expériences 2 et 3) et aussi lorsque la visibilité en vision centrale est moins bonne (Expérience 5). Mais si on considère la taille globale de l'effet d'interférence périphérique, sans prendre en compte les variables de similarité physique ou de visibilité, cette taille est généralement plus petite que celle de l'effet d'interférence des BFs sur la catégorisation des HFs en vision centrale mesurée dans l'Expérience 1, surtout pour les temps de réponse (%change $\simeq 2\%$ pour l'effet de la vision périphérique contre $\simeq 4\%$ pour l'effet des BFs en vision centrale). On pourrait alors faire l'hypothèse que le poids des prédictions issues de la vision périphérique relativement à celles issues des BFs en vision centrale pourrait être légèrement inférieur. Cependant, l'interférence en vision centrale a été testée grâce à des stimuli hybrides qui ne sont pas écologiques. De nouveaux paradigmes expérimentaux permettant de comparer directement le poids de l'influence prédictive issue des traitements réalisés en vision centrale et celui de l'influence prédictive issue des traitements réalisés en vision périphérique doivent être développés pour pouvoir répondre à cette question. Une perspective intéressante serait de comparer l'effet d'interférence induit par les BFs d'un objet en vision centrale à celui induit par les BFs contextuelles périphériques sur la catégorisation des HFs de l'objet (Figure 4.1).

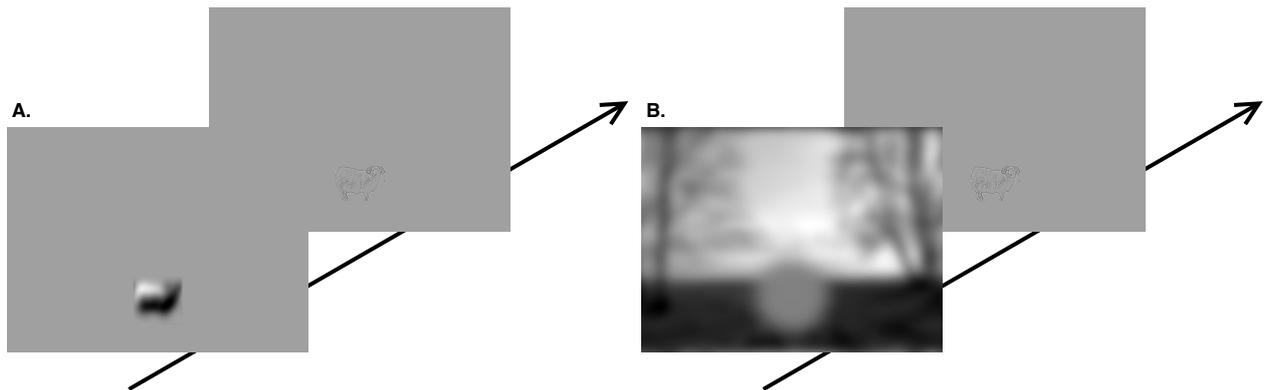


Figure 4.1: Protocole expérimental pouvant évaluer le poids relatif des prédictions issues du traitement des BFs de l'objet en vision centrale (A.) et de celles issues du traitement des BFs du contexte en vision périphérique (B.) sur la catégorisation des HFs de l'objet en vision centrale.

Indépendamment de cette question, les tailles d'effet sur les temps de réponses correctes des Expériences 2 et 3 indiquent que la présence d'une information congruente en vision périphérique (par rapport à une information incongruente) ferait gagner une vingtaine de millisecondes pour reconnaître la catégorie sur-ordonnée d'une scène en vision centrale, si tant est qu'il y a une similarité des informations physiques portées par la vision centrale et la vision périphérique (i.e. en condition PHYSIM). Les tailles d'effet des Expériences 4 et 5 indiquent que la présence d'un contexte congruent (par rapport à un contexte incongruent) ferait gagner une dizaine de millisecondes pour reconnaître

la catégorie sur-ordonnée d'un objet en vision centrale. Ainsi, les tailles de l'effet d'interférence périphérique sont plutôt petites, de l'ordre d'une quinzaine de millisecondes. Cependant, un temps si petit n'est pas anodin si on considère la rapidité avec laquelle le système visuel traite l'information en comparaison avec d'autres modalités perceptives.

Un enjeu en psychologie est de pouvoir juger de l'importance pratique des effets mesurés. En contexte naturel, est-il avantageux de pouvoir bénéficier d'une quinzaine de millisecondes grâce à l'intégration de prédictions issues de la vision périphérique ? Notre point de vue est que le contexte expérimental ne mime pas du tout une situation dans laquelle ce genre de mécanismes peut être bénéfique, la tâche n'impliquant aucun enjeu pour le participant. En situation naturelle, on peut au contraire penser qu'un si petit effet peut avoir des conséquences importantes. Par exemple, on peut imaginer qu'un conducteur reconnaîtra plus rapidement un enfant qui traverse soudainement devant son véhicule s'il conduit dans une rue qui longe une école (situation congruente) que s'il conduit sur une route départementale en pleine campagne (situation incongruente). Dans ce genre de situation, même quelques dizaines de millisecondes peuvent être importantes pour être en mesure de réagir rapidement.

Par ailleurs, il faut ici à nouveau souligner que nous avons utilisé des tâches de catégorisation sur-ordonnée. L'effet aurait pu être plus important en utilisant une tâche de catégorisation au niveau de base ou sous-ordonné. Ces niveaux de catégorisation sont accessibles plus tardivement, laissant plus de temps pour que des processus prédictifs puissent influencer le traitement visuel. De manière générale, il est tout à fait probable que l'influence de l'information disponible en vision périphérique varie considérablement en fonction de la situation. On peut penser intuitivement que le système cognitif a tout intérêt à utiliser l'information en vision périphérique pour accéder à une représentation rudimentaire latente du monde visuel car elle est tout d'abord utile à la reconnaissance rapide, mais elle permet aussi d'anticiper sur le traitement du signal entrant en vision centrale, afin de répondre le plus vite et le plus efficacement possible à des situations qui pourraient provoquer une erreur de prédiction.

1.3. Limites méthodologiques

La rédaction de ce chapitre est l'occasion de revenir sur la méthodologie utilisée dans nos travaux. Nous pouvons identifier plusieurs limites méthodologiques. D'abord, bien que le contrôle expérimental de la position des stimuli dans le champ visuel soit au centre de nos recherches, nous ne disposons pas au moment des expérimentations d'un appareil de mesures oculaires pour nous assurer que les participants fixent bien le stimulus à catégoriser en vision centrale. Il faut toutefois noter que les tâches de catégorisation demandées au participant dans les expériences de ce travail de thèse paraissent difficiles ou coûteuses à réaliser sans maintenir la fixation. Par exemple, dans les Expériences 4 et 5, étant donné que la tâche était de fixer et catégoriser l'objet, il aurait été inefficace et même coûteux de faire des saccades vers d'autres endroits de la scène sans perturber la réalisation de la tâche de catégorisation. En effet, les saccades sont généralement

initiées en 100-150 ms (Fischer & Weber 1993). Si les participants ont fait une saccade dans la scène (présentée pendant 150 ms dans la condition SOA-150ms) avant de faire une nouvelle saccade vers l'objet (également présentée pendant 150 ms), cela aurait posé des contraintes sévères sur (i) le temps restant pour traiter l'information visuelle dans la scène ou (ii) le temps restant pour initier une nouvelle saccade vers l'objet et le catégoriser. D'ailleurs, la proportion de réponses correctes était élevée (dans l'Expérience 4 et dans l'Expérience 5 lorsque la visibilité de l'objet était raisonnable), ce qui laisse penser que les mouvements oculaires n'ont pas gêné la catégorisation des objets.

Ensuite, on ne peut pas exclure que le paradigme d'interférence puisse induire des biais de demande chez les participants. En effet, bien que le participant ne doit pas prêter attention à l'information en vision périphérique pendant les tâches, il semble raisonnable de penser qu'il puisse se rendre compte que nous avons manipulé la congruence. Ainsi, il est possible que certains participants aient compris l'objectif de notre recherche, modifiant alors leur comportement. Ceci est d'autant plus probable chez des étudiants et étudiantes en psychologie. Cependant, il nous semble que les mesures du temps de réponses correctes et du seuil de visibilité (Expérience 5) sont plutôt résistantes aux biais de demande. La proportion de réponses correctes a pu être en revanche plus impactée.

Concernant spécifiquement la proportion de réponses correctes, nous avons remarqué que chez les sujets sains (en particulier dans les Expériences 2, 3 et 4), cette proportion était souvent très haute (nos proportions moyennes allant de 0.93 à 0.97) et traduisait parfois un effet plafond. Travailler sur une échelle de mesure aussi limitée peut être très restrictif et notamment limiter la possibilité de détecter des effets (Wagenmakers, Kryptos, Criss, & Iverson 2012). Par ailleurs, comme nous l'avons souligné plus haut, cet effet plafond qui traduit la facilité de nos tâches peut expliquer les effets de petite taille observés sur les temps de réponses correctes (en d'autres termes, l'effet d'interférence aurait pu être plus grand avec une tâche plus difficile).

Enfin, un facteur expérimental qui a pu réduire l'effet d'interférence périphérique est la présentation d'un masque après le stimulus (*backward masking*) pour contrôler les effets de persistance rétinienne. En effet, dans une étude récente, Tang et al. (2018) ont montré que l'utilisation d'un masque pouvait réduire les influences *feedback*. Dans cette étude les participants devaient catégoriser des objets qui étaient soit intacts, soit partiellement visibles (donc plus difficilement reconnaissables). Un masque (un bruit distribué en $1/f$ comme celui utilisé dans certaines de nos expériences) était présenté après l'objet, avec un SOA variant entre 25 et 150 ms. Les résultats ont montré que plus le SOA entre l'objet et le masque était court et moins les performances étaient bonnes, et ce surtout pour les objets partiellement visibles. Les auteurs ont alors suggéré que les mécanismes prédictifs *feedback* qui permettent la complétion de pattern nécessaire pour reconnaître les objets partiellement visibles étaient perturbés par la présence d'un masque. L'utilisation d'un masque dans les paradigmes d'interférence doit donc être questionnée. Il est d'ailleurs intéressant de remarquer que nous n'avons pas observé d'effet d'interférence

lorsque les stimuli central et périphérique étaient présentés simultanément dans les expériences dans lesquelles nous avons utilisé un masque (i.e. Expérience 4), alors que c'était le cas dans les expériences où nous n'avons pas utilisé de masque (i.e. Expériences 2 et 3) - bien que comme nous l'avons déjà évoqué, cette divergence de résultat peut aussi être expliquée par divers autres facteurs, comme par exemple le niveau de difficulté de la tâche (la catégorisation d'objet dans les Expériences 4 et 5 pouvant être [encore] plus facile que la catégorisation des scènes dans les Expériences 2 et 3).

1.4. Apports des résultats pour les théories prédictives de la reconnaissance visuelle

1.4.1. Arguments empiriques d'une influence prédictive du contenu périphérique

Dans toutes nos expériences, nous avons observé un effet d'interférence de l'information périphérique sur le traitement de l'information centrale, l'effet étant de faible magnitude mais récurrent. Comme nous l'avons vu dans la partie théorique de cette thèse, les effets d'interférence peuvent être expliqués par deux hypothèses (qui avaient été formulées initialement par [Henderson & Hollingworth 1999](#) pour les effets de contexte en général). La première suppose que les informations du champ visuel central et du champ visuel périphérique sont traitées de manière principalement indépendante, et en parallèle. Sous cette hypothèse, l'effet d'interférence résulterait d'une mise en concurrence des deux traitements réalisés en parallèle de façon purement ascendante. La seconde, qui correspond au modèle FP-peri, suppose qu'il y a une précedence de traitement de l'information périphérique (le *gist*) et que le résultat de son traitement est rétro-injecté sur le système de reconnaissance afin de guider le traitement de l'information ascendante en HFs provenant de la rétine centrale.

Les Expériences 2 et 3 nous permettaient difficilement de départager clairement ces deux hypothèses puisque la présentation des scènes centrale et périphérique était simultanée. Dans l'Expérience 4, nous avons présenté les stimuli central et périphérique de façon séquentielle en manipulant la pré-exposition de la scène périphérique, i.e. en faisant varier le SOA entre la scène périphérique et l'objet à catégoriser. L'idée sous-jacente est que si les représentations de la scène périphérique et de l'objet interagissent seulement lors d'un traitement ascendant parallèle, l'effet d'interférence devrait être le même quelle que soit la durée pendant laquelle la scène était perçue en amont. Au contraire, un modèle linéaire a montré que l'effet d'interférence de la scène périphérique sur la reconnaissance de l'objet augmentait avec le temps de pré-exposition de la scène périphérique. Ce résultat ne nous semble alors compatible qu'avec l'hypothèse d'une influence de mécanismes prédictifs issus du traitement de la vision périphérique.

Par ailleurs, dans l'Expérience 5, nous avons demandé aux participants de catégoriser des objets plus ou moins visibles. Nous avons observé que la congruence de l'information

en vision périphérique diminuait le niveau de visibilité nécessaire pour pouvoir reconnaître correctement l'objet en vision centrale. Une qualité insuffisante des informations sensorielles ascendantes en vision centrale pourrait donc être compensée par des prédictions pertinentes issues de la vision périphérique. En ce sens, ce résultat soutient le modèle FP-péri dans lequel un système de reconnaissance intègre l'information ascendante et les prédictions provenant notamment du *gist* extrait en vision périphérique.

1.4.2. Nature et mécanismes des prédictions

Un enjeu des modèles prédictifs de la reconnaissance visuelle est de s'accorder sur la nature des prédictions et le niveau du traitement visuel sur lequel les mécanismes prédictifs s'exercent. Nous avons également identifié ces deux problématiques dans le cadre du modèle FP-péri que nous avons résumées dans la Figure 4.2.

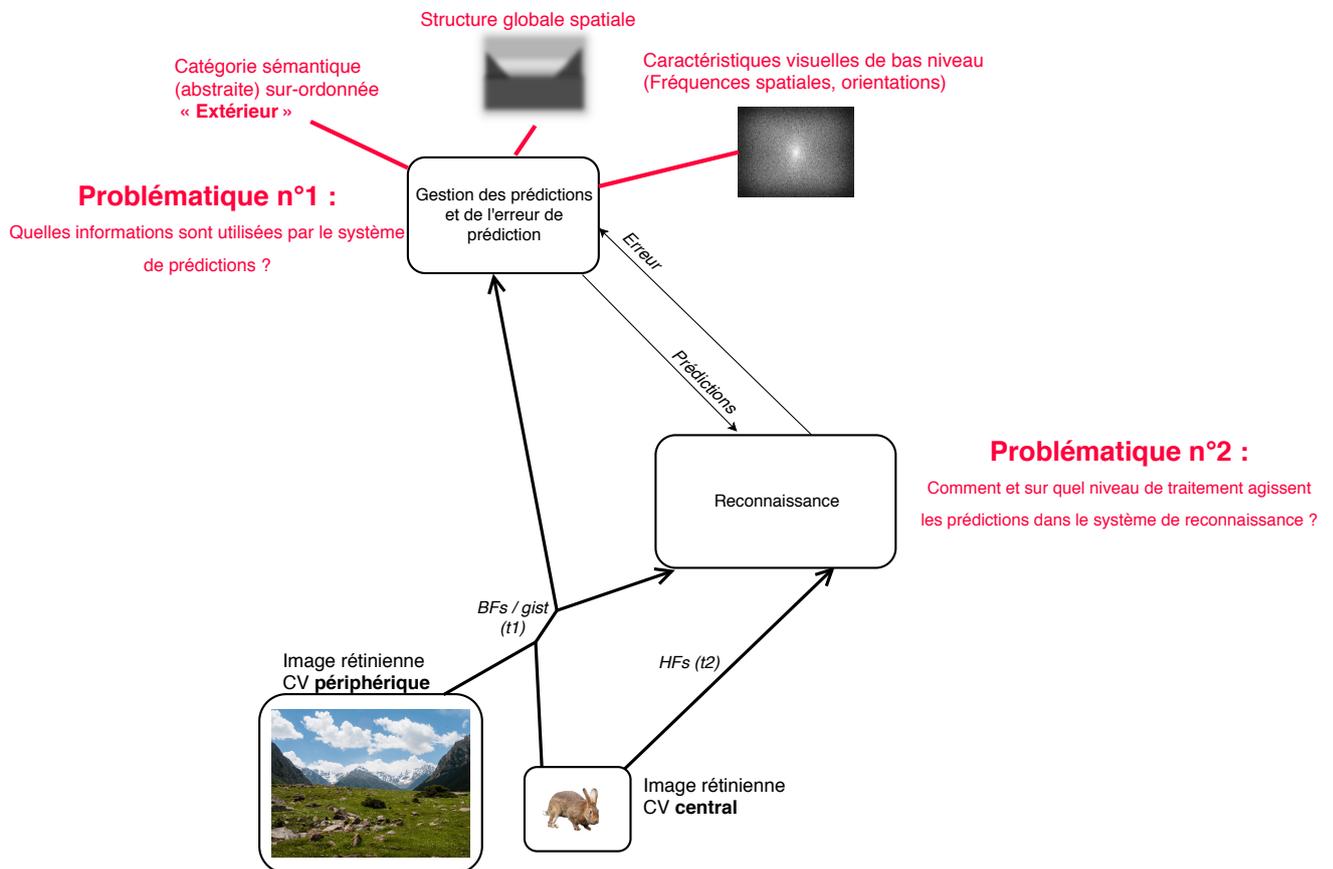


Figure 4.2: Deux problématiques à traiter pour une meilleure compréhension des mécanismes prédictifs dans un modèle FP-péri.

La première problématique est de déterminer quelle est l'information qui est utilisée par le système de gestion des prédictions. Nous avons postulé que c'est le *gist* de la scène qui est utilisé, et plus spécifiquement les BF. Cependant, l'enjeu est de savoir à quel niveau de l'émergence du *gist* les prédictions commencent à être déclenchées, et donc quel type de contenu est utilisé pour former les prédictions. Est-ce que le système

de prédiction utilise une information sémantisée de haut niveau (contenu conceptuel du *gist*), indépendante des caractéristiques visuelles de bas niveau ? Ou est-ce que le système de prédiction utilise des caractéristiques visuelles de bas niveau (contenu perceptif du *gist*) ? Et dans ce cas, est-ce une information contenue dans le spectre d'amplitude ? Ou est-ce plutôt une information spatiale telle que la configuration de la scène décrite par l'agencement des blobs de luminance ?

Les résultats des expériences du Chapitre 2 peuvent apporter quelques éléments de réponse à cette première problématique. D'abord, nous avons manipulé la similarité physique entre les scènes centrale et périphérique dans les Expériences 2 et 3. Les scènes étaient soit physiquement similaires, c'est-à-dire que leur configuration spatiale (corrélation IP) et leur spectre d'amplitude (corrélation SA) étaient similaires, soit physiquement dissimilaires. Le raisonnement était que si le système de prédiction utilise une information physique de bas niveau, alors la scène périphérique devrait influencer d'autant plus la scène centrale lorsque leurs propriétés physiques sont similaires. Dans ces deux expériences, les résultats ont montré une interférence de la scène périphérique seulement lorsque les propriétés physiques des scènes étaient similaires, suggérant que le système de reconnaissance aurait tendance à intégrer en un même percept des informations visuelles centrales et périphériques qui ont les mêmes caractéristiques dans l'espace de Fourier et/ou la même configuration spatiale. Ce résultat est en faveur de l'hypothèse d'une utilisation par le système de prédiction des caractéristiques visuelles de bas niveau disponibles en vision périphérique. On peut alors se demander si l'accès à une catégorie sémantique (i.e. au *gist* conceptuel, par exemple la catégorie sur-ordonnée de la scène) est absolument nécessaire pour que le système puisse générer un signal prédictif.

Ainsi, afin de déterminer si la seule information physique de bas niveau peut suffire à déclencher des prédictions, ou si au contraire le système de prédiction code aussi une information sémantique, nous avons manipulé la structure de phase de la scène périphérique dans l'Expérience 5. Chaque scène contextuelle en vision périphérique était présentée dans sa version intacte et dans sa version *scrambled*. La procédure de brouillage (*scrambling*) de phase préserve l'information contenue dans le spectre d'amplitude des scènes intactes, c'est-à-dire le contenu en orientation et en fréquences spatiales, mais perturbe la phase de l'image. Ainsi dans les scènes *scrambled*, on perd la configuration spatiale de la scène et donc la possibilité d'en extraire une information sémantique. Les résultats ont montré que les scènes *scrambled* n'interféraient pas sur la catégorisation de l'objet, indiquant que la seule information contenue dans le spectre d'amplitude ne suffit pas à déclencher des prédictions. A priori, ce résultat suggère qu'une représentation sémantique/abstraite des scènes est nécessaire pour déclencher des prédictions. A posteriori, cette dernière interprétation doit tout de même être formulée avec précaution. Nous avons en effet évoqué le fait qu'en plus d'empêcher l'émergence d'une représentation sémantique, la méthode du *scrambling* de phase perturbait la configuration spatiale de l'information visuelle. Cette méthode ne nous permet donc pas de supprimer l'information sémantique sans supprimer en même temps certaines caractéristiques de bas

niveau. Ceci nous amène à questionner la pertinence d'utiliser le *scrambling* de phase pour étudier le rôle des caractéristiques visuelles de bas niveau dans l'activation de prédictions. Nous avons donc proposé qu'avec les scènes intactes, les prédictions pouvaient être issues non pas nécessairement d'une représentation sémantique, mais d'une représentation de la configuration spatiale ou de l'agencement spatial des informations visuelles de la scène. A l'appui de cette nouvelle hypothèse, nous avons observé **(i)** un effet d'interférence dans les Expériences 2 et 3 dans le cas particulier où les scènes centrale et périphérique étaient physiquement similaires à la fois sur leur spectre d'amplitude et leur configuration spatiale, et **(ii)** que cette interférence disparaît lorsqu'on conserve le spectre d'amplitude, mais qu'on supprime la configuration spatiale dans l'Expérience 5.

La seconde problématique concerne le niveau de traitement sur lequel les prédictions agissent dans le système de reconnaissance. Une idée intuitive est que les prédictions opèrent de manière rétinotopique. Autrement dit, un blob de luminance à un endroit donné de la scène est prédictif des HFs qui se situent au même endroit sur l'image rétinienne (BFs et HFS corrélées ou superposées). Pour reprendre l'exemple donné par Petras et al. (2019), percevoir la forme d'un chien (BFs) limite l'endroit où la fourrure (HFs) peut se trouver. Dans ce cas, cela implique que le système de prédiction renvoie le signal prédictif sous un format rétinotopique sur les bas niveaux du traitement visuel. Néanmoins, dans cette thèse, nous avons implicitement fait l'hypothèse d'un type de mécanismes prédictifs qui ne seraient pas rétinotopiques. Avec le modèle FP-péri, nous postulons en effet que le signal BFs extrait en vision périphérique peut prédire et aider à traiter les détails en vision centrale. Comme nous l'avons évoqué dans le Chapitre 1 (sous partie 3.2), l'idée de l'existence de prédictions non rétinotopiques est soutenue par les travaux de Williams et al. (2008) qui ont montré que des aires visuelles rétinotopiques dédiées à la vision centrale peuvent coder l'information visuelle périphérique. L'effet d'interférence périphérique mesuré dans nos expériences suggère que des mécanismes prédictifs non rétinotopiques peuvent effectivement avoir lieu.

Dans le cadre des théories du codage prédictif, on pourrait envisager que les mécanismes prédictifs exercent une influence dès les premiers niveaux dans la hiérarchie du traitement visuel (Friston 2009, 2010; Kok & de Lange 2015; Rao & Ballard 1999), facilitant ainsi l'analyse des caractéristiques de bas niveau. Mais on peut aussi imaginer que de tels mécanismes, puisqu'ils sont non rétinotopiques, s'exercent seulement sur les hauts niveaux du traitement visuel. Par exemple, saisir rapidement le contexte en vision périphérique (e.g., un magasin de meuble) pourrait permettre de pré-activer dans le système de reconnaissance des représentations de haut niveau concernant l'objet (e.g., concernant les tables, les fauteuils, les buffets, etc.), c'est-à-dire des représentations invariantes à la taille, à la position, à la luminosité, etc., sans influencer les traitements perceptifs qui ont permis leur émergence. À un niveau cortical, ces pré-activations pourraient se manifester dans le cortex inférotemporal ou même un peu plus en amont dans la hiérarchie de traitement, par exemple dans la partie du LOC se trouvant dans le gyrus fusiforme. En effet, les populations de neurones du LOC et du cortex inférotemporal

contiennent des représentations d'objet invariantes aux conditions visuelles (taille rétinienne, position, luminosité, etc. Grill-Spector et al. 2001; Kourtzi & Kanwisher 2001; Vogels & Orban 1996). Cette idée est compatible avec le modèle de Bar (2003) et l'étude de Bar et al. (2006) qui suggère que le cortex orbitofrontal - considéré par ces auteurs comme le site cortical d'un système de prédiction - exerce une influence sur le gyrus fusiforme lors de la catégorisation d'objets. Similairement, les analyses de connectivité fonctionnelles appliquées dans l'étude en IRMf de Kauffmann et al. (2015c) ont montré que l'effet d'interférence des BFs sur les HFs de scènes hybrides s'associait à une connectivité fonctionnelle entre le cortex orbitofrontal et le gyrus fusiforme. Les résultats de l'Expérience 5 peuvent nous apporter un élément de réponse sur le niveau de traitement sur lequel s'exercent les prédictions. En effet, si les prédictions s'exerçaient seulement à partir d'un haut niveau de traitement, il faudrait déjà que la construction d'une représentation haut niveau de l'objet soit possible sur la base de traitements seulement ascendants, notamment en représentant la forme de l'objet (par extraction et complétion des contours). Hors, dans notre Expérience 5, les stimuli objets étaient très bruités. La Figure 4.3 illustre bien le fait qu'il est impossible de former une représentation de haut niveau par le biais de seuls mécanismes ascendants pour les niveaux de visibilité autour du seuil (i.e. niveaux 3 et 4). Pourtant, non seulement ces stimuli pouvaient être correctement catégorisés, mais nous avons en plus observé un effet d'interférence. Cette observation est plutôt en faveur de mécanismes prédictifs qui s'exercent dès les bas niveau du traitement visuel. Nous faisons l'hypothèse que dans l'Expérience 5, les mécanismes prédictifs ont plutôt facilité des mécanismes de complétion de contours, dans des régions possiblement aussi précoces que V1 (Kok et al. 2016; Kok, Failing, & de Lange 2014; Kok, Jehee, & De Lange 2012). Bien entendu, cela n'exclut pas la possibilité que les mécanismes prédictifs agissent aussi sur les traitements de plus haut niveau dans le système de reconnaissance. La puissance des théories du codage prédictif réside justement dans la proposition que les mécanismes de codage prédictif interviennent à chaque étape du traitement de l'information, ce qui comme le souligne Clark (2013) permet de s'affranchir d'une conception partitionnée de la perception et de la cognition.

Comment alors un signal prédictif non-rétinotopique pourrait agir à si bas niveau du traitement ? Il est par exemple possible que l'information codée dans le système de prédiction (e.g., la configuration spatiale de la scène et sa catégorie sur-ordonnée) puisse être associée à la probabilité de rencontrer tel ou tel type de caractéristique élémentaire du signal visuel. Les prédictions pourraient alors permettre de favoriser la détection de ces caractéristiques de très bas niveau au niveau d'aires visuelles comme V1. Par exemple, la perception rapide de la configuration spatiale d'une scène de plage (e.g., trois gros blobs arrangés horizontalement correspondant à la plage, à la mer et au ciel) pourrait être retransformée afin de préactiver les populations de neurones de V1 répondant aux orientations horizontales (afin de mieux traiter la ligne d'horizon par exemple) et celles répondant aux fréquences moyennes et hautes (afin de mieux traiter les grains de sable et les traces d'ondulations dans le sable par exemple; voir des exemples de scènes et leurs spectres d'amplitude moyen sur la Figure 4.4). À l'inverse, lors d'une

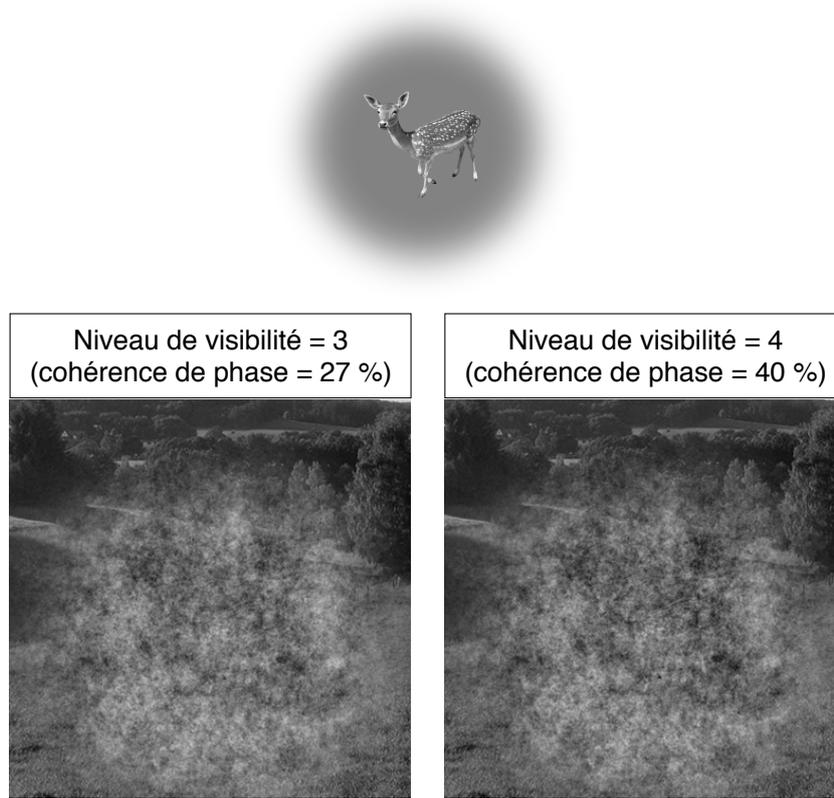


Figure 4.3: Exemple d'un objet de la catégorie 'animaux' (un faon) utilisé dans l'Expérience 5. Le seuil de catégorisation correcte se situait entre les niveaux de visibilité 3 (27 % de cohérence de phase) et 4 (40 % de cohérence de phase).

promenade en forêt, la perception rapide de la configuration spatiale de la scène (e.g., des gros blobs en hauteur à cause du feuillage des arbres et une forme triangulaire en bas formée par le point de fuite du chemin) pourrait permettre de préactiver les populations de neurones de V1 répondant aux HFs en hauteur (feuillage des arbres) et celles répondant aux orientations verticales de moyenne ou haute fréquence spatiale (alignement des troncs d'arbres). Les exemples proposés sont arbitraires, et il est possible que ce soit au contraire la perception des orientations obliques qui soit renforcée par ces mécanismes de prédictions, si l'on considère les travaux qui démontrent une sensibilité accrue aux orientations obliques par rapport aux orientations cardinales dans le cadre de la perception d'images aux caractéristiques spectrales naturelles (i.e. distribution $\frac{1}{f^\alpha}$ de l'énergie ; [Essock, DeFord, Hansen, & Sinai 2003](#); [Hansen & Essock 2006](#)). Dans tous les cas, cette hypothèse pourrait être testée avec nos stimuli en présentant d'abord le contexte (les scènes utilisées dans les Expériences 4 et 5), puis des stimuli sans signification variant sur les caractéristiques de bas niveau (e.g. orientations verticales de basse fréquence, contours courbés en HFs...). Si une information contextuelle pertinente en vision périphérique permet une meilleure perception de ces caractéristiques, le seuil de contraste pour détecter ces caractéristiques (e.g., stimuli de type gabors de différentes

orientations et fréquences spatiales, voir [Teufel et al. 2018](#) pour une procédure similaire) devrait être plus bas avec un contexte congruent (e.g., une scène de forêt → des orientations verticales).

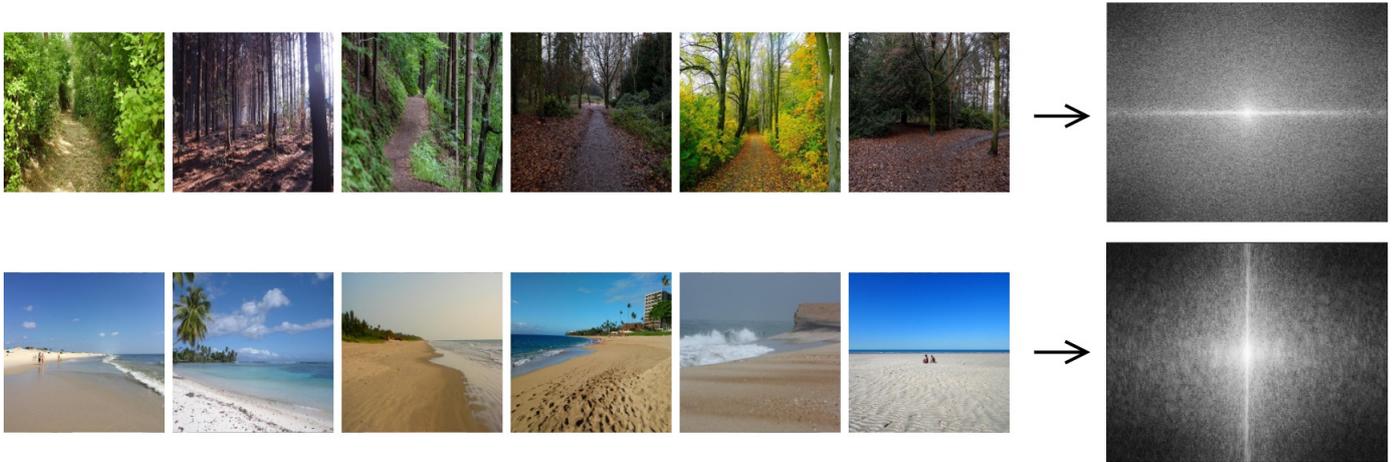


Figure 4.4: Deux types de contextes pouvant donner lieu à deux types de prédictions de ‘bas niveau’. En haut : scènes de forêts. Le spectre d’amplitude moyen montre qu’il y a beaucoup d’énergie dans les hautes voire très hautes fréquences (feuillage), et dans les orientations verticales (troncs des arbres). En bas : scènes de plages. Le spectre d’amplitude moyen montre qu’il y a beaucoup d’énergie dans les moyennes et hautes fréquences (grains de sables et traces d’ondulations dans le sable), et dans les orientations horizontales (ligne d’horizon).

En conclusion, au regard de l’ensemble des informations apportées par les Expériences du Chapitre 2 et des discussions précédentes, la Figure 4.5 permet de préciser le modèle FP-péri. D’abord, on peut envisager que le système de prédiction utilise en priorité une information de plutôt bas niveau, possiblement en interaction avec une première catégorisation sémantique sur-ordonnée. Concernant les informations de bas niveau en question, après avoir éliminé un effet possible d’une seule information contenue dans le spectre d’amplitude, nous pensons que la configuration spatiale de la scène serait la plus susceptible de jouer ce rôle. Les prédictions se projettent ensuite sur le système de reconnaissance, où elles modèleraient le traitement de l’information détaillée provenant du champ visuel central dès les premiers niveaux de traitement. Cela n’exclut pas que les prédictions puissent aussi agir sur des niveaux de traitement plus hauts (flèches en pointillés sur la figure).

1.4.3. Le cortex orbitofrontal comme site cortical d’un système de prédiction

L’idée que le cortex orbitofrontal puisse être le site cortical d’un module prédictif a d’abord été proposée par [Bar \(2003\)](#). Un argument pour cette hypothèse est que le cortex orbitofrontal est fortement connecté au cortex inférotemporal chez le macaque ([Cavada et al. 2000](#)). Par ailleurs, des études lésionnelles montrent que cette région est impliquée

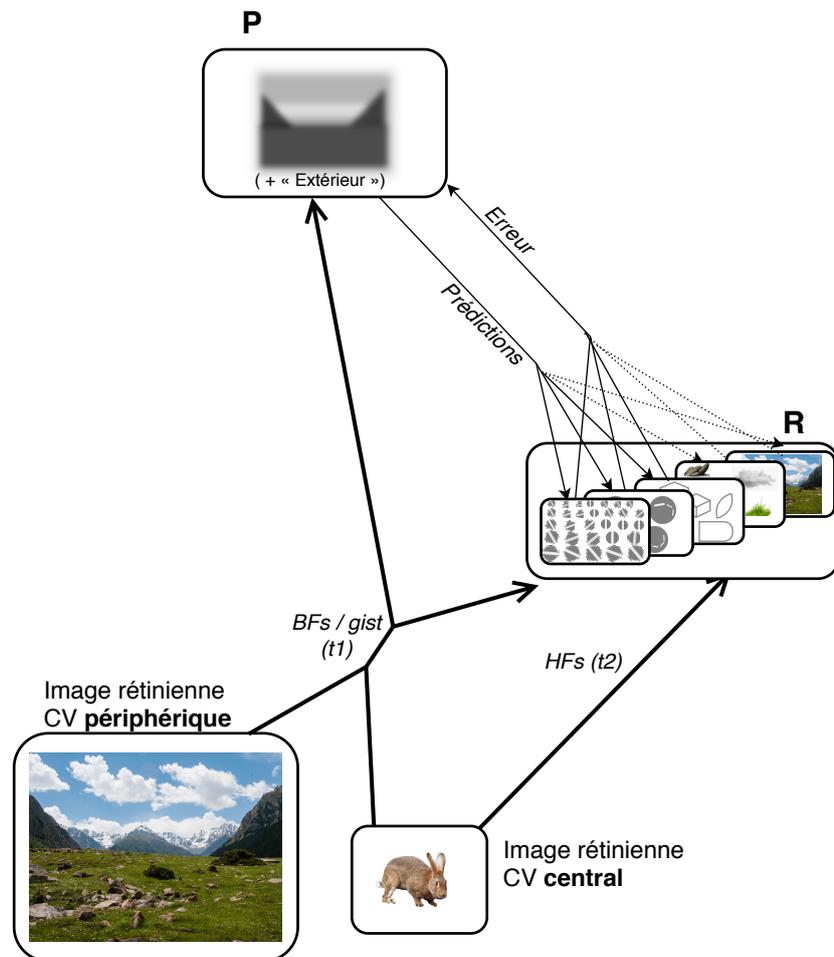


Figure 4.5: Précision du modèle FP-péri en fonction des deux problématiques définies dans la Figure 4.2.

dans le traitement des informations visuelles (Thorpe et al. 1983). Dans le cadre de la perception de scènes, l'étude IRMf de Kauffmann et al. (2015c) a montré que l'effet d'interférence des BFes sur les HFS de scènes hybrides s'accompagnait d'une activation de cette région. Nous avons nous aussi observé une activation du cortex orbitofrontal dans notre étude réalisée sur les sujets sains (Expérience 3), ainsi que dans celle réalisée auprès de patients glaucomateux (Expérience 7). Dans l'Expérience 3, nous avons observé un résultat très similaire à celui de Kauffmann et al. (2015c) puisque l'effet d'interférence de la scène périphérique s'accompagnait d'une activation orbitofrontale dont le pic avait des coordonnées très proches de celles relevées par ces auteurs (et également proches de celles relevées par Bar et al. 2006). Le cortex orbitofrontal pourrait alors être impliqué dans le codage prédictif de l'information visuelle que celle-ci provienne du champ visuel central (Kauffmann et al. 2015c) ou périphérique (Expérience 3). Ce résultat étaye l'hypothèse d'un mécanisme commun de codage prédictif du signal provenant de la rétine centrale et de la rétine périphérique, comme nous le postulons a priori avec le modèle FP-péri. Au seuil d'activation que nous avons défini pour l'analyse des résultats en IRMf, nous n'avons observé cette activation que lorsque les deux scènes étaient physiquement similaires,

suggérant que le cortex orbitofrontal serait particulièrement sensible aux caractéristiques physiques des stimuli. À noter que [Frey et Petrides \(2000\)](#) ont montré que le cortex orbitofrontal était impliqué dans une tâche de rétention en mémoire de patterns visuels sans signification (les participants devaient essayer d'inscrire en mémoire les patterns physiques des stimuli et ne pouvaient pas s'appuyer sur des stratégies sémantiques), ce qui suggère que cette région a priori non visuelle peut coder des informations visuelles.

Dans l'Expérience 7, les résultats montrent que le réseau cérébral impliqué dans la tâche de catégorisation de scènes faiblement contrastées chez les participants contrôles comprenait (1) le cortex occipital postérieur (2) le cortex occipito-temporal ventral (gyrus fusiforme, PPA), et (3) le cortex frontal inférieur, notamment dans sa partie orbitofrontale, ainsi que l'insula antérieure. Au contraire, les régions occipito-temporales étaient moins recrutées chez les patients. Par ailleurs, le cortex orbitofrontal/insulaire n'était pas activé chez deux des patients testés, et à l'inverse sur-engagé chez la troisième patiente testée. Cette étude exploratoire laisse penser que le glaucome pourrait causer un désengagement progressif de cette région, possiblement lié à un dysfonctionnement du système de prédictions, ou au contraire un sur-engagement de cette région pour pallier au déficit visuel. Cette proposition n'est pour le moment qu'hypothétique et de nouvelles études sont nécessaires pour étudier plus spécifiquement le fonctionnement de cette région dans le glaucome.

Enfin, nous avons observé de manière inattendue que les activations orbitofrontales associées à l'effet d'interférence dans l'Expérience 3 s'étendaient jusqu'au cortex insulaire antérieur. De plus, lors de la tâche de catégorisation de scènes de l'Expérience 7, les participants contrôles activaient soit le cortex orbitofrontal, soit l'insula antérieure. Bien que nous ne nous attendions pas à observer une activation insulaire dans nos études, ces résultats pourraient suggérer que le cortex orbitofrontal et l'insula antérieure supportent un ensemble de mécanismes communs. Le caractère multimodal de ces régions pourrait indiquer qu'elles ont un rôle dans la gestion des mécanismes de prédictions pour n'importe quelle modalité perceptive, mais aussi dans la gestion des prédictions pour des processus cognitifs de plus haut niveau comme par exemple la cognition sociale ou la prise de décision. Dans une étude en IRMf récemment parue, [Loued-Khenissi, Pfeuffer, Einhäuser, et Preuschoff \(2020\)](#) ont étudié le réseau cérébral qui sous-tend les mécanismes prédictifs de manière générale. Leur approche est de placer les participants dans des situations expérimentales d'incertitude perceptive (avec une tâche de report de percept d'un stimulus bi-stable) et décisionnelle (avec une tâche de prise de décision financière). Selon les auteurs, les processus de prédiction et de gestion de l'erreur de prédiction sont inextricablement liés à tout état d'incertitude, qu'elle soit d'origine perceptive ou décisionnelle. Leurs résultats ont montré que l'insula antérieure était particulièrement activée lorsque ces tâches induisaient une incertitude liée à l'erreur de prédiction. Les auteurs ont alors proposé que l'insula aurait alors un rôle dans la gestion de l'erreur de prédiction dans des situations ambiguës. D'autres études ont mis en évidence une implication de l'insula antérieure dans la résolution des erreurs de prédictions lors de la

prise de risques (de nouveau évaluée grâce à des tâches de prise de décision financière; [Bossaerts 2010](#); [Preuschoff, Quartz, & Bossaerts 2008](#)). Cette région corticale pourrait alors avoir un rôle de gestion des erreurs de prédiction de manière globale.

De manière générale, nous pouvons conclure qu'un nombre assez important d'études, y compris l'Expérience 3 et dans une moindre mesure l'Expérience 7 de cette thèse, suggèrent que le cortex orbitofrontal (et possiblement l'insula) pourrait avoir un rôle dans le codage prédictif ou le pilotage des mécanismes de codage prédictif. Cependant, le rôle exact de cette région dans le traitement de l'information visuelle est loin d'être clair. En effet, si l'on veut pouvoir considérer cette région comme le site de l'implémentation corticale du système de gestion des prédictions du modèle FP-péri (ou de tout autre modèle prédictif de la reconnaissance visuelle), il faut pouvoir déterminer quel type d'information est codée dans cette région, et quel est le rôle exact de ses interactions fonctionnelles avec le cortex occipito-temporal ([Bar et al. 2006](#); [Kauffmann et al. 2015c](#)). Par ailleurs, le cortex orbitofrontal est une région intégrative complexe jouant un rôle dans de nombreuses fonctions cognitives (e.g., l'intégration des stimuli gustatifs et olfactifs, [Kringelbach & Rolls 2004](#); [Rolls 2019](#), l'apprentissage opérant et la prise de décision, [Rudebeck & Rich 2018](#), le traitement de stimuli affectifs, [Rolls 2019](#), ou encore la cognition sociale, [Beer, John, Scabini, & Knight 2006](#)). Il ne doit donc pas être exclu que l'implication du cortex orbitofrontal dans les études utilisant le paradigme d'interférence ([Kauffmann et al. 2015c](#), et Expérience 3 de cette thèse) ou dans les études sur la catégorisation d'objets ([Bar et al. 2006](#)) puisse être liée à une fonction cognitive tout autre que celle d'un rôle de gestion des prédictions. Une méta-analyse de [Nee, Wager, et Jonides \(2007\)](#) a examiné les activations observées dans des études en neuroimagerie de tâches impliquant la résolution d'une interférence/d'un conflit (e.g., Stroop, go/no-go, stop signal, effet Simon). Les résultats de la méta-analyse ont montré que le cortex frontal inférieur et l'insula antérieure étaient des régions fréquemment impliquées dans ces tâches qui impliquent la résolution d'un conflit ou d'une interférence. Les auteurs suggèrent plus spécifiquement que l'activation de ces régions serait liée à la sélection de la réponse pendant ces tâches par le biais de mécanismes d'inhibition ou de facilitation. Il est alors possible que les activations orbitofrontales et insulaires associées à l'interférence périphérique que nous avons observées dans l'Expérience 3 correspondent à un processus post-perceptif de sélection de la réponse. À noter cependant que ces régions étaient actives chez nos quatre participants contrôles de l'Expérience 7, alors que la tâche de catégorisation utilisée dans cette étude n'impliquait pas d'interférence.

1.4.4. Un système prédictif moins efficace lorsque la vision périphérique est dégradée

Un autre apport de ces travaux de thèse concerne l'approche pathologique que nous avons adoptée pour étudier le rôle de la vision périphérique dans la reconnaissance visuelle. Nous avons abordé cette problématique en étudiant les capacités de reconnaissance visuelle de personnes atteintes de glaucome. L'intérêt du glaucome est que la vision

périphérique est déficitaire à long terme et de manière irréversible. On peut donc envisager que la maladie cause une désorganisation progressive de tout mécanisme reposant sur le traitement des informations visuelles périphériques. Ainsi, et puisque l'hypothèse de cette thèse est que les mécanismes de prédiction dans la reconnaissance visuelle reposent en partie sur le traitement des informations périphériques, nous avons proposé que le glaucome devrait causer une perturbation immuable du système de prédiction. L'Expérience 6 nous a permis de démontrer que les patients atteints de glaucome avaient des difficultés à reconnaître des scènes en vision centrale lorsque le contraste est bas (mimant une situation de très faible luminosité). En comparant la sensibilité de patients glaucomateux à une tâche de détection, impliquant seulement des processus de très bas niveau, à une tâche de catégorisation, impliquant des processus de bas mais aussi de haut niveau menant à la reconnaissance visuelle, nous avons constaté que le coût des processus de haut niveau était plus élevé chez les patients que chez les participants contrôles, et ce même pour les patients qui n'avaient pas de déficit sensoriel dans le champ visuel central. La reconnaissance visuelle des patients en vision centrale serait donc moins efficace à cause d'un dysfonctionnement du système de prédiction. L'Expérience 7 pourrait suggérer que ce dysfonctionnement se traduirait au niveau cortical par un désengagement des régions impliquées dans la reconnaissance visuelle de scènes, comme la PPA ou le cortex orbitofrontal. Nous pensons alors que le glaucome constitue un modèle pathologique intéressant pour une approche de l'étude de la cognition visuelle qui prendrait en compte la vision périphérique.

2. Perspectives théoriques et implications cliniques

2.1. Un modèle basé sur la perception du *gist* doit-il être absolument un modèle fréquentiel ?

Le modèle FP-péri dérive du modèle FP et il est donc par définition un modèle fréquentiel. Dans les deux modèles, on suppose que c'est le contenu en BFs qui contribue à l'émergence du *gist* et à l'activation rapide de prédictions qui guident le traitement des HFs. Mais dans les expériences de cette thèse (mise à part l'Expérience 1), nous n'avons pas filtré les stimuli. Nous sommes partis du principe que la rétine périphérique agit comme un filtre passe-bas naturel, à cause des grands champs récepteurs des cellules ganglionnaires de cette partie de la rétine. Par ailleurs, cela permettait de dénaturer au minimum le stimulus afin de conserver les conditions les plus écologiques possibles. Cependant, l'absence de filtrage nous empêche de conclure que c'est spécifiquement et seulement le traitement des BFs qui contribue au *gist* et à l'activation rapide de prédictions. Nous l'avons vu dans le Chapitre 1, la perception en vision périphérique n'est pas seulement caractérisée par la mauvaise acuité (i.e. l'impossibilité d'extraire des HFs), mais aussi par le phénomène d'encombrement (*crowding*). Le *gist* d'une scène extrait en vision périphérique ne serait alors pas seulement caractérisé par une information globale en BFs, mais aussi par des percepts d'ensemble dérivés de mécanismes de *summary statistics*.

Il est intéressant de noter qu'en dehors du champ de recherche de la reconnaissance visuelle, les modèles de la vision périphérique sont en fait très peu concernés par une approche fréquentielle du traitement visuel. Ces modèles ont comme point commun de reposer sur le phénomène de *crowding* plutôt que sur celui de la baisse d'acuité. Par exemple, Ruth Rosenholtz et ses collaborateurs ont proposé le modèle TTM (*Texture Tiling Model*; Balas, Nakano, & Rosenholtz 2009; Ehinger & Rosenholtz 2016; Rosenholtz, Huang, & Ehinger 2012). Ce modèle propose que la vision périphérique représente l'image sous forme de patchs (tuiles) texturés, l'équivalent dans l'expérience subjective de *summary statistics* calculés localement sur des larges régions du champ visuel (de plus en plus larges avec l'excentricité). Ce type de modèle peut assez bien rendre compte des performances pour reconnaître des stimuli encombrés en vision périphérique. Par exemple, Balas et al. (2009) ont utilisé un algorithme de synthèse de texture (Portilla & Simoncelli 2000) pour synthétiser de nouvelles images ayant approximativement les mêmes statistiques que des images de base. Les images synthétisées par ce procédé sont appelées des *mongrels* (Figure 4.6). Les auteurs faisaient l'hypothèse que le mongrel est une visualisation approximative de l'information encodée lorsque l'image intacte est perçue en vision périphérique. Ils ont comparé ensuite les performances à différentes tâches (e.g., une tâche d'identification de lettres) entre un groupe de participants qui

réalisent la tâche sur les mongrels en vision centrale, et un autre groupe avec les stimuli intacts en vision périphérique. Leurs résultats ont montré que les performances à ces différentes tâches sur les stimuli mongrels présentés en vision centrale corrélaient bien avec les performances à ces mêmes tâches utilisant les stimuli intacts présentés en vision périphérique.

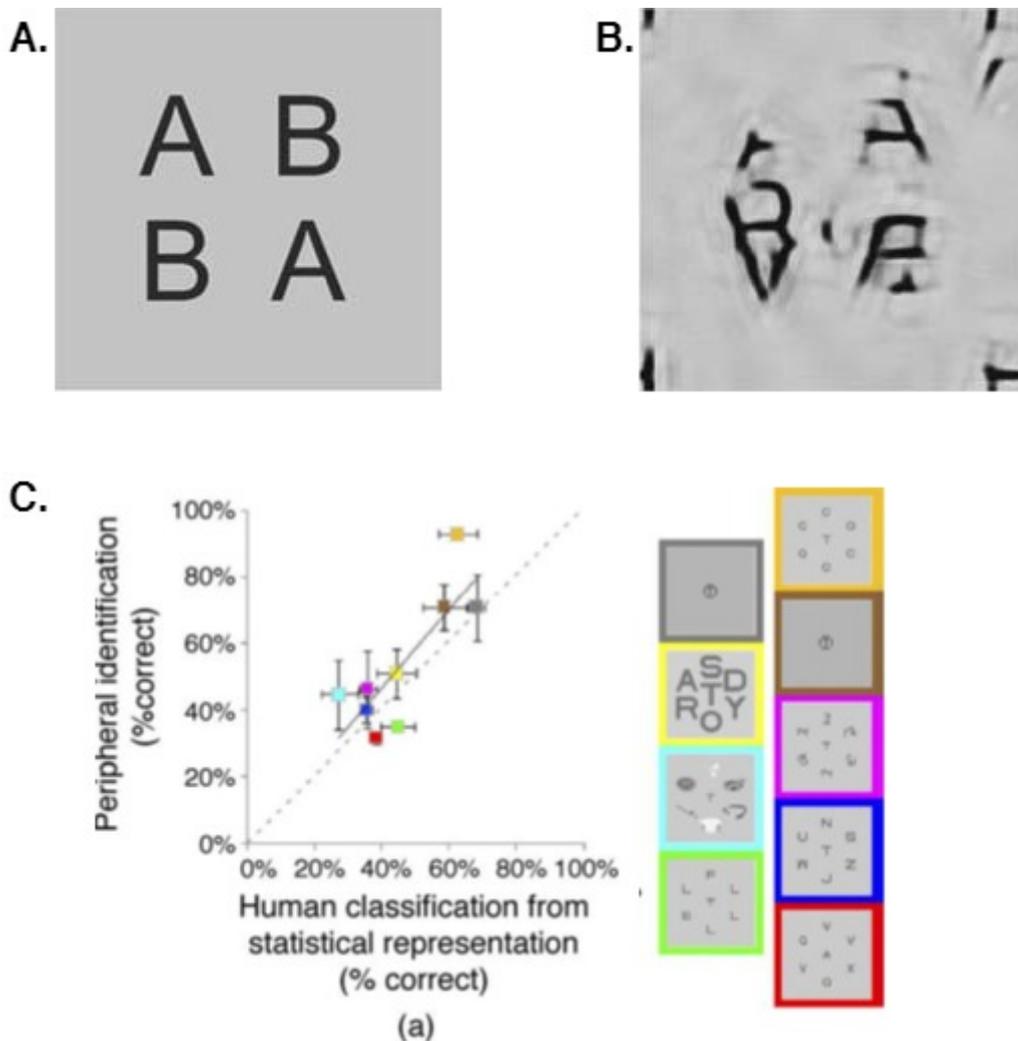


Figure 4.6: Stimuli et résultats de l'étude de Balas et al. (2009). **A.** Stimulus intact. **B.** 'Mongrel' synthétisé à partir du stimulus intact avec l'algorithme de synthèse de textures de Portilla et Simoncelli (2000). **C.** Résultats de l'expérience : Les performances à différentes tâches sur les stimuli mongrels (axe X) présentés en vision centrale corrélaient bien avec les performances à ces mêmes tâches utilisant les stimuli intacts présentés en vision périphérique (axe Y).

Par ailleurs, en utilisant le même procédé expérimental, cette même équipe a montré que le modèle TTM prédit assez bien les performances à une tâche de catégorisation du *gist* de scènes (Ehinger & Rosenholtz 2016). Dans cette étude, un premier groupe de participants devait catégoriser des scènes intactes présentées en vision périphérique (la partie fovéale de la scène était masquée ; Figure 4.7-A). Un second groupe de participants

devait catégoriser des scènes mongrels (Figure 4.7-B) qui étaient synthétisées à l'aide de l'algorithme de Portilla et Simoncelli (2000) en créant des patches de textures superposés respectant des statistiques calculées localement sur des *pooling regions* de taille augmentant avec l'excentricité (suivant la loi de Bouma 1970). Les résultats ont montré que les performances à la tâche de catégorisation de scènes intactes en vision périphérique corrélaient bien avec les performances à la tâche de catégorisation de scènes mongrels. Les auteures ont aussi testé les performances pour catégoriser des versions des scènes floutées avec un flou gaussien qui simulaient la perte d'acuité visuelle avec l'excentricité (Figure 4.7-C). Les performances pour catégoriser les scènes intactes corrélaient aussi plutôt bien avec les performances pour catégoriser les scènes floues, mais moins qu'avec les performances pour catégoriser les scènes mongrels synthétisées avec le TTM (scènes floues : $R^2 = 0.62$, scènes mongrels : $R^2 = 0.75$). Même si ces scènes floues n'ont pas été construites en réalisant un filtrage fréquentiel (i.e. ce ne sont pas des BFs *per se*), ce résultat peut indiquer qu'un modèle de vision périphérique basé sur les *summary statistics* pourrait mieux expliquer la capacité de reconnaissance du *gist* de scène qu'un modèle basé sur les BFs.

Une autre manière d'aborder la problématique du type de contenu perceptif extrait en vision périphérique est d'adopter une approche phénoménologique. En effet, lorsqu'on y prête attention, on peut avoir l'impression que l'apparence de notre environnement en vision périphérique ne correspond pas exactement à celle d'une image filtrée en BFs. Au delà du flou et des formes rudimentaires, l'apparence en périphérie prend aussi un aspect texturé et un peu chaotique avec une confusion entre les différents éléments, qui correspondrait plutôt à un stimulus encombré comme les mongrels de Balas et al. (2009, 4.6-B) qu'à une image filtrée en BFs. Les modèles de la vision périphérique reposant sur les *summary statistics* comme le TTM peuvent-ils alors rendre compte de l'apparence qu'ont les stimuli plus complexes que sont les scènes visuelles en vision périphérique? Cette question a été étudiée grâce à une approche basée sur le report de l'expérience subjective du sujet. Plus exactement, cette approche consiste à essayer de générer des stimuli métamériques, c'est-à-dire des stimuli différents physiquement mais dont l'apparence est la même (i.e. indifférenciables du point de vue de l'expérience subjective). Freeman et Simoncelli (2011) ont testé un type de stimuli métamériques construits en remplaçant les différentes parties périphériques de scènes (*pooling regions*) par des patches de textures qui tiennent compte des *summary statistics* locales des différentes régions de la scène. Les textures étaient calculées sur des régions de plus en plus grandes en périphérie pour tenir compte de l'agrandissement des champs récepteurs avec l'excentricité rétinienne (le procédé est similaire à celui utilisé par Ehinger & Rosenholtz 2016). Ces stimuli ont été présentés à des participants dans les 26° centraux de leur champ visuel, et les résultats ont montré que deux stimuli différents construits selon ce procédé étaient indifférenciables. Ce résultat est toutefois modéré par une récente étude de Wallis et al. (2019) qui montre que si les participants doivent comparer une image modifiée de la même manière que celles de Freeman et Simoncelli (2011) et l'image intacte de base, ces deux images sont facilement discriminables par les participants. À ce jour (et à

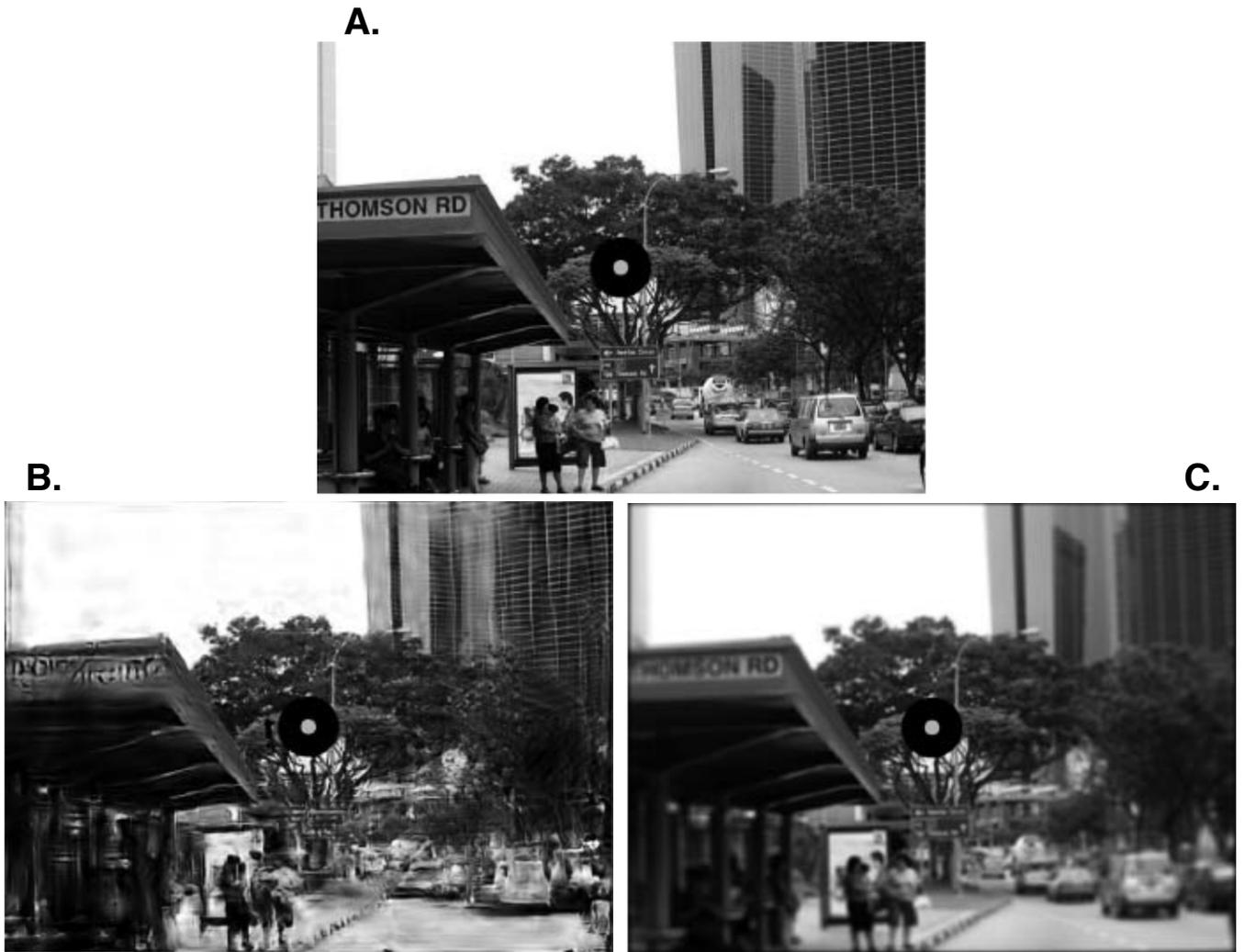


Figure 4.7: Stimuli de l'étude de Ehinger et Rosenholtz (2016).
A. Scène intacte. **B.** Scène mongrel. **C.** Scène floutée avec un flou gaussien augmentant avec l'excentricité.

notre connaissance), il n'existe donc pas de véritables images métamériques qui simulent l'apparence en vision périphérique. Il est intéressant de noter que les auteurs concluent que l'impossibilité de créer de vraies images métamériques en considérant seulement les *summary statistics* peut être liée au fait que les modèles de texture perturbent la segmentation globale de la scène, autrement dit l'extraction de la configuration spatiale globale de la scène, qui est en revanche bien préservée dans les BFs. Cette remarque des auteurs peut être mise en lien avec nos différents résultats des expériences du Chapitre 2 qui suggèrent que la configuration spatiale de la scène est une information importante dans le *gist* et qu'elle serait utilisée pour activer des prédictions.

Ces travaux sur les modèles de type 'texture' suggèrent qu'un bon modèle de la vision périphérique doit considérer les *summary statistics*, mais aussi des caractéristiques du signal permettant de préserver la segmentation globale, comme par exemple les BFs. Mettre en relation ces deux champs de recherche permettrait de mieux spécifier quels

sont les contenus visuels qui permettent de saisir le *gist* d'une scène, ce qui a toute son importance dans un modèle prédictif de la reconnaissance visuelle qui considère la vision périphérique.

2.2. Modes de vision, modèle de hiérarchie inverse et implications pour les personnes atteintes d'un glaucome

Les travaux réalisés au cours de cette thèse concernaient la reconnaissance visuelle, une problématique située plutôt dans une conception passive de la perception visuelle. En effet, nous l'avons souligné au début de ce document, percevoir et reconnaître visuellement notre environnement semble automatique et ne demande aucun effort. Il suffit d'ouvrir les yeux et en une fixation, nous pouvons déterminer le lieu dans lequel nous nous trouvons et reconnaître grossièrement les objets qui nous entourent, cette reconnaissance ultra-rapide reposant sur l'extraction du *gist* de la scène. Cependant, en conditions naturelles, la vision est bien souvent un comportement actif. Autrement dit, les modèles de reconnaissance visuelle 'fixation unique' doivent pouvoir être complétés par une approche active de la perception visuelle, qui considère le fait que le comportement visuel est dynamique, sans cesse guidé par les objectifs du sujet. Il est intéressant de noter qu'un modèle de reconnaissance visuelle prédictif qui inclut la vision périphérique, comme le modèle FP-péri, peut tout à fait être considéré dans une approche de vision active. En effet, l'avantage d'accumuler rapidement de l'information en vision périphérique est aussi de pouvoir guider les mouvements oculaires, en offrant une prévisualisation grossière en périphérie. Cette prévisualisation déclenche un signal prédictif qui pourra être intégré aux détails apportés par la fixation suivante.

Dans ce contexte écologique où peuvent être considérées à la fois la perception active et la perception passive, certaines propositions théoriques peuvent être utiles pour pouvoir mieux comprendre le rôle des différents contenus visuel (*gist*, BFs, HFs, ...) dans les modes de perception visuelle. Le référentiel proposé par Hochstein et Ahissar (2002) est intéressant à cet égard. Dans leur modèle de hiérarchie inverse, ces auteurs ont proposé que la structure bidirectionnelle d'un cortex visuel hiérarchiquement organisé permet de sous-tendre deux modes d'analyse visuelle, distincts mais complémentaires. Les premiers traitements ascendants tout le long de la voie ventrale seraient majoritairement inconscients / implicites. Leur rapidité permettrait d'aboutir à une première représentation consciente / explicite, du *gist* de la scène via l'activation d'aires de relativement haut niveau. Ensuite, l'accès à des représentations plus précises serait possible par une analyse procédant de manière *top-down* le long de la hiérarchie corticale pour intégrer l'information plus détaillée. Ces deux modes de perception consciente sont résumées par les auteurs par deux concepts : **(i)** *vision at a glance* dans un premier temps, qui permettrait l'accès au *gist* de la scène, avec une attention divisée sur l'ensemble de la scène qui requiert les champs récepteurs larges d'aires de haut niveau, et **(ii)** *vision with scrutiny*

dans un second temps, qui repose sur des mécanismes d'attention focalisée et permet l'analyse des détails (Figure 4.8).

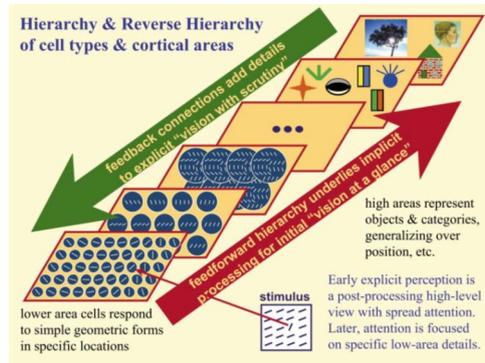


Figure 4.8: Modèle de hiérarchie inverse de Hochstein et Ahissar (2002).

Un mode de vision basée sur une représentation latente du *gist* (*at a glance*) peut être suffisante dans bien des situations où la vision est ‘passive’ - et ceci même pendant une échelle de temps plus longue pendant laquelle les mouvements oculaires interviennent. Bon nombre des tâches sensori-motrices que nous réalisons au quotidien sont guidées par la vision mais ne nécessitent pas un mode de vision contrôlé et précis. Par exemple, lorsqu’on se lève un matin de semaine, on peut passer en revue les tâches qui nous attendent dans la journée tout en enchaînant une série de comportements ‘zombies’ (sortir une tasse du placard, préparer la machine à café, ranger quelques objets, faire couler le café, etc.). Ces tâches ne nécessitent pas une attention visuelle très focale, ni un traitement minutieux des détails de la scène visuelle. Il en est de même par exemple pour faire un trajet piéton routinier, comme par exemple pour se rendre sur son lieu de travail. Au contraire, chercher ses clés demande une analyse minutieuse des différentes régions de la scène. La lecture est une activité qui repose aussi particulièrement sur la possibilité d’extraire les HFs, donc sur une analyse minutieuse. Nous avons nous-même proposé que l’énorme effet d’interférence des HFs sur la catégorisation des scènes en BFs puisse être du à une stratégie d’analyse minutieuse de type vision *with scrutiny*. Ainsi, un mode de vision *with scrutiny* reposerait plutôt sur une attention focalisée au niveau de la vision centrale alors qu’un mode de vision basé sur le *gist* reposerait plutôt sur une attention distribuée sur l’ensemble du champ visuel, donc sur la vision périphérique. Hochstein et Ahissar (2002) font aussi remarquer que les deux modes de vision seraient étroitement liés à deux modes de recherche visuelle. Un mode de vision reposant sur le *gist* est associé à une stratégie de recherche visuelle en parallèle (de type ‘pop-out’), qui serait rapide, automatique et peu coûteuse. Au contraire, un mode de vision *with scrutiny* est associé à la recherche visuelle sérielle, qui elle serait longue et coûteuse (Treisman & Gelade 1980). Ainsi, le modèle de hiérarchie inverse de Hochstein et Ahissar (2002) implique qu’un mode de perception visuelle *with scrutiny* exigerait des processus supplémentaires, engageant les connexions descendantes et donc des mécanismes plus importants de codage prédictif, par rapport à un mode de vision basé sur la perception

latente du *gist*. En d'autres termes, le mode *with scrutiny* engagerait un coût cognitif plus important. Par ailleurs, il est intéressant de noter que chacun de ces modes cause des 'erreurs d'attention' différentes. Dans un mode de vision reposant sur le *gist*, la sensibilité au détail est basse et des changements dans la scène, même importants, peuvent ne pas être perçus consciemment (notamment lorsqu'ils surviennent pendant une saccade) tant qu'ils n'affectent pas la structure du *gist*. On appelle ce phénomène la cécité au changement (Rensink, O'Regan, & Clark 1997). À l'inverse, un mode de vision *with scrutiny* serait plutôt lié à un autre type d'erreur d'attention. Puisque l'attention est focalisée sur un endroit de la scène, des phénomènes de cécité d'inattention ont de grandes chances de se produire. La cécité d'inattention intervient dans le cas où nous sommes focalisés sur une information particulière de la scène visuelle et qu'on ne remarque pas un élément qui apparaît ou qui change dans la scène, même très près du point de fixation. C'est par exemple le cas dans la célèbre expérience du gorille invisible (Simons & Chabris 1999) qui témoigne qu'il est possible de regarder quasiment directement un objet, et néanmoins de continuer à ne pas le percevoir consciemment.

Ainsi la vision reposerait sur un système à deux régimes. Un premier mode basé sur le *gist* et caractérisé par une attention distribuée sur l'ensemble du champ visuel permettrait le guidage des comportements automatiques et la recherche visuelle "en parallèle" (de type pop-out), mais ne permettrait pas un accès aux détails. Ce mode serait peu coûteux et peut donner lieu à un phénomène de cécité au changement. Un second mode *with scrutiny*, caractérisé par une attention focale centrée sur la vision centrale, permettrait l'analyse minutieuse des détails et la recherche visuelle "sérielle" lorsque la cible est difficilement trouvable (chercher ses clés). Le mode *with scrutiny* est plus coûteux et peut donner lieu à des cécités d'inattention. L'être humain basculerait constamment entre ces deux modes afin d'apporter à la conscience les contenus dont il a besoin pour réaliser la tâche en cours. Cette conception passive/active de la perception visuelle peut avoir des implications importantes pour les patients atteints de glaucome. En effet, l'extraction du *gist* des scènes chez ces patients est mise à mal par la perte de vision périphérique. Nous avons aussi proposé que l'atteinte progressive et irréversible de la vision périphérique cause des dysfonctionnements de tout processus reposant sur la vision périphérique. Il est alors possible que les patients glaucomateux puissent moins s'appuyer sur un mode de vision 'passif' basé sur le *gist*. Ces patients basculeraient alors principalement sur un mode de vision *with scrutiny*, sur-engageant au quotidien la vision centrale et l'analyse minutieuse des détails et diminuant le poids de leur vision périphérique résiduelle. Ceci pourrait avoir une implication forte dans la vie quotidienne des patients puisque le coût cognitif associé à ce mode est plus important. Par ailleurs, rappelons que même le mode *with scrutiny* serait indirectement impacté par le glaucome puisque comme nous l'avons vu, les mécanismes de reconnaissance visuelle seraient moins efficaces chez les patients à cause d'un dysfonctionnement du système prédictif. Ces changements fonctionnels pourraient expliquer les nombreuses difficultés dont les patients font l'expérience, notamment dans ce qui concerne les comportements 'zombies'. Par exemple, l'atteinte et la saisie d'objets sont perturbées (Koticha et al.

2009) et les patients expriment souvent des difficultés lors des déplacements à pieds et pour réaliser des activités de la vie quotidienne comme cuisiner ou faire le ménage (Nelson et al. 1999), des comportements dont on peut penser qu'ils reposent principalement sur un mode de vision passive reposant sur le *gist*. La recherche visuelle est également impactée (N. D. Smith et al. 2011, 2012). Par ailleurs, le basculement principalement sur un mode *with scrutiny* pourrait provoquer une importante fatigue visuelle. B. L. Lee et al. (1998) ont montré que les patients reportent sept fois plus des sensations de fatigue voire de douleurs oculaires que des sujets sains. Dans une étude qualitative, Glen et Crabb (2015) ont mené des entretiens semi-dirigés auprès de 16 patients atteints de glaucome afin de mieux comprendre les difficultés concrètes vécues par les patients dans leur vie quotidienne et les éventuelles stratégies d'adaptation qui pouvaient être mises en place. Ces entretiens révèlent des points intéressants concernant le mode de vision que certains patients utilisent. Voici des extraits de ces entretiens, traduits de l'anglais. Les patients racontent leur expérience concernant la conduite automobile :

« C'est vrai que j'ai l'impression que je tourne volontairement ma tête à gauche et à droite plus souvent. Je veux dire plus que je ne le faisais auparavant, parce qu'il faut que je fasse particulièrement attention. »

« Les gens me disaient souvent [quand je conduisais encore] : "n'importe qui penserait que tu es en train de passer ton permis, la manière que tu as de constamment tourner la tête". »

« Je suis très consciente que lorsque je conduis, même lorsqu'il fait beau, je ne fais que bouger les yeux sans cesse [...]. »

ou encore à propos des déplacements à pieds :

« Quand je monte ou je descends un escalier, je bouge ma tête vers le haut et vers le bas sans cesse [...]. »

« Traverser la route [est encore un autre exemple]. Tu dois regarder et regarder encore pour être sûr que rien n'arrive. »

Ainsi, évaluer les modes de perception des patients atteints de glaucome pourrait être une piste intéressante pour mieux comprendre le trouble glaucomateux. Cette évaluation pourrait se faire en comparant les patients à des contrôles appariés en âge sur une batterie de tâches impliquant la vision basée sur le *gist*. On peut prédire que les patients seraient moins efficaces pour catégoriser très rapidement le *gist* (pour des scènes présentées sur l'ensemble du champ visuel), pour calculer des estimateurs moyens dans des tâches de *summary statistics* comme celle utilisée par Ariely (2001), ou encore pour trouver automatiquement ('pop-out') une cible dans un paradigme de recherche visuelle parallèle. On peut aussi prédire que les patients seraient moins sujets à la cécité au changement, et plus sujets à la cécité d'inattention, que des participants contrôles. La difficulté méthodologique (de taille...) sera que les stimuli doivent être présentés dans l'ensemble du champ visuel tout en respectant absolument la vision résiduelle propre à

chaque patient pour que le déficit sensoriel seul ne puisse pas expliquer de moins bonnes performances des patients sur des tâches de *gist*.

S'il est possible de démontrer que les patients témoignent de difficultés pour ces tâches basées sur un mode de vision *gist*, on peut penser qu'une forme de rééducation peut être mise en place, par exemple en entraînant les patients sur ces tâches. À cet égard, il est intéressant de noter qu'il est assez bien établi, dans le domaine de l'apprentissage perceptif, que les tâches faciles comme la recherche visuelle parallèle sont particulièrement sujettes à un apprentissage perceptif très rapide et stable, et que cet apprentissage peut se généraliser à travers les stimuli (voir par exemple [Ahissar & Hochstein 1997](#)) ce qui pourrait être un avantage considérable pour la perspective clinique que nous proposons. Dans un tout autre contexte, des travaux ont montré que l'entraînement à des jeux vidéos d'action, ou leur pratique régulière, pouvaient entraîner de meilleures aptitudes à répartir l'attention sur l'ensemble du champ visuel (i.e. basculer en mode *gist*; [Achtman, Green, & Bavelier 2008](#); [Green & Bavelier 2003](#); [Spence & Feng 2010](#)). Ainsi, on peut penser que le cadre de l'apprentissage perceptif pourrait permettre aux patients de réinvestir de la confiance dans leur vision périphérique résiduelle et ainsi de pouvoir 'redémarrer' progressivement les mécanismes qui reposent sur le *gist*, c'est-à-dire à la fois un mode de vision *gist* peu coûteux, mais aussi les mécanismes prédictifs qui permettent de mieux analyser les détails.

Annexe 1

How does information from low and high spatial frequencies interact during scene categorization?

Louise Kauffmann^{a,b,†}, Alexia Roux-Sibilon^{a,†}, Brice Beffara^a, Martial Mermillod^a, Nathalie Guyader^c and Carole Peyrin^a

^aDepartment of Psychology, University of Grenoble Alpes, CNRS, LPNC UMR 5105, Grenoble, France; ^bNeural Mechanisms of Human Communication Research Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ^cImage and Signal Department, University of Grenoble Alpes, GIPSA-lab UMR5216, Grenoble, France

ABSTRACT

Current models of visual perception suggest that, during scene categorization, low spatial frequencies (LSF) are rapidly processed and activate plausible interpretations of visual input. This coarse analysis would be used to guide subsequent processing of high spatial frequencies (HSF). The present study aimed to further examine how information from LSF and HSF interact and influence each other during scene categorization. In a first experimental session, participants had to categorize LSF and HSF filtered scenes belonging to two different semantic categories (artificial vs. natural). In a second experimental session, we used hybrid scenes as stimuli made by combining LSF and HSF from two different scenes which were semantically similar or dissimilar. Half of the participants categorized LSF scenes in hybrids, and the other half categorized HSF scenes in hybrids. Stimuli were presented for 30 or 100 ms. Session 1 results showed better performance for LSF than HSF scene categorization. Session 2 scene categorization was faster when participants attended and categorized LSF than HSF scene in hybrids. The semantic interference of a semantically dissimilar HSF scene on LSF scene categorization was greater than the semantic interference of a semantically dissimilar LSF scene on HSF scene categorization, irrespective of exposure duration. These results suggest a LSF advantage for scene categorization, and highlight the prominent role of HSF information when there is uncertainty about the visual stimulus, in order to disentangle between alternative interpretations.

ARTICLE HISTORY

Received 19 December 2016
Accepted 20 June 2017

KEYWORDS

Spatial frequency;
coarse-to-fine; hybrids;
semantic interference

Introduction

At a glance, the human visual system is able to robustly process and categorize complex stimuli such as natural scenes (e.g., Thorpe, Fize, & Marlot, 1996), despite their infinite variability in terms of objects they contained, or their spatial configuration. Converging data from neurophysiological recordings in primates (Bullier, 2001; De Valois, Albrecht, & Thorell, 1982; Poggio, 1972; Shams & von der Malsburg, 2002; Shapley & Lennie, 1985; Van Essen & Deyoe, 1995) and psychophysical studies in human (Hughes, Nozawa, & Kitterle, 1996; Parker, Lishman, & Hughes, 1996; Schyns & Oliva, 1994; for a review see Sowden & Schyns, 2006) indicate that the visual system

extract visual information through a set of channels/filters differently tuned to specific orientations and spatial frequency bands of the visual input. Based on these data, current models of visual perception have emphasized the role of spatial frequency information for visual categorization (Bar, 2003; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Peyrin et al., 2010). According to these models, visual analysis begins with the parallel extraction of different elementary features at different points of the spatial frequency spectrum that provide different information about the visual scene. Lower spatial frequencies (LSF) provide coarse information, such as the global shape and structure of a visual scene, and are

CONTACT Louise Kauffmann  kauffmann@cbs.mpg.de  Department of Psychology, University of Grenoble Alpes, CNRS, LPNC UMR 5105, F-38040 Grenoble, France; Neural Mechanisms of Human Communication Research Group, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany

[†]These authors contributed equally.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/13506285.2017.1347590>.

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

predominantly conveyed through fast magnocellular channels. Higher spatial frequencies (HSF) provide finer information about the scene, such as edges or object details, and are conveyed more slowly through parvocellular channels. On the basis of the neurophysiological properties of the magno- and parvocellular pathways (Bullier, 2001; Maunsell et al., 1999) and results of psychophysical studies in humans (Hughes et al., 1996; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994), it has been suggested that visual scene analysis follows a predominantly coarse-to-fine processing sequence. LSF information would be extracted first, allowing a coarse parsing of the visual input, prior to the analysis of fine information contained in HSF. In this theoretical framework, it was also hypothesized that rapid LSF information may guide the subsequent processing of HSF (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Trapp & Bar, 2015).

A predominant coarse-to-fine processing of scenes was first evidenced in a behavioural study by Schyns and Oliva (1994). The stimuli used by these authors were hybrid scenes made by adding a scene filtered in LSF (e.g., a highway) to another scene filtered in HSF (e.g., a city) and presented for 30 or 150 ms. They showed that, for short exposure (30 ms), categorical perception of hybrids was based on the category of the LSF scene which illustrates a perception dominated by the scene in LSF. However, for longer exposure (150 ms), the perception of hybrids was dominated by the scene in HSF. Importantly, participants were not aware that the hybrids contained two different scene categories. These results therefore suggested that, in the absence of awareness of any apparent conflict between two scenes in the stimulus, LSF information is predominantly used at early stages of visual processing whereas HSF are used at later stages in accordance with a coarse-to-fine visual processing sequence. Further studies supported these results by showing for example that scenes filtered in LSF are categorized more quickly than scenes filtered in HSF (e.g., Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015) and that the processing of LSF before HSF information is more advantageous for scene categorization than the reverse processing sequence (i.e., HSF before LSF; e.g., Kauffmann, Chauvin, et al., 2015; Musel et al., 2012). However, little is known about the integration of LSF and HSF

before the full recognition of scenes and the potential influence of LSF information on HSF processing.

Recent neuroimaging studies investigated that issue using hybrid scenes as stimuli and examining directly how the categorization of the HSF component in hybrids is influenced by the LSF component (Kauffmann, Bourgin, Guyader, & Peyrin, 2015; Mu & Li, 2013). In these studies, participants were aware of the two different scenes in hybrids and were instructed to ignore the LSF scene in hybrids, attend the HSF scene, and categorize the latter. The two scenes composing the hybrids were either semantically similar (e.g., a natural scene in LSF superimposed with a natural scene in HSF) or dissimilar (e.g., a natural scene in LSF superimposed with a man-made scene in HSF). The categorization of HSF scenes was impaired when the two scenes composing the hybrids were semantically dissimilar. Although it was not relevant to perform the task, the semantic information contained in the LSF scene was still processed and interfered with the categorization of the HSF scene when the two scenes were semantically dissimilar. This semantic interference effect was even greater when the two scenes composing the two semantically dissimilar scenes were physically similar (Mu & Li, 2013), suggesting that the spatial overlap between LSF and HSF caused greater impairment of HSF categorization in the absence of any semantic congruence between the two scenes. The semantic interference effect was thus interpreted as the signature of the integration of semantic information from both LSF and HSF and the influence of LSF over HSF scene categorization.

On a neurobiological level, results from ERP recordings (Mu & Li, 2013) revealed that the semantic interference effect was associated with a negative frontal component (N1) 122 ms after stimulus onset, suggesting that this short amount of time is sufficient for the integration of semantic information from both LSF and HSF in frontal areas. Results from fMRI (Kauffmann, Bourgin, et al., 2015) further showed that the semantic interference effect involved the inferior frontal gyrus (including the orbitofrontal cortex) and the inferotemporal cortex, and was associated with greater functional connectivity between these two regions. Overall, results from these two studies supported the proactive model of visual recognition (Bar, 2003, 2007; Trapp & Bar, 2015). In that theoretical context, the semantic interference effect in frontal and inferotemporal areas was interpreted as reflecting

erroneous predictions generated in the inferior frontal cortex based on LSF information of hybrids in the semantically dissimilar condition that led to impaired HSF scene categorization in the inferotemporal cortex, resulting in greater connectivity between these areas.

These studies provided supplementary argument in favour of a coarse-to-fine categorization of scenes, by showing that fast processing of LSF information is automatic and cannot be voluntarily inhibited. Furthermore, they suggested that the extraction of semantic information contained in LSF strongly influences the processing of HSF information during scene categorization. However, whether HSF processing can also influence LSF categorization was not the main focus of these studies. Although the prominent role of LSF information for rapid scene categorization has been extensively investigated and documented (Hughes et al., 1996; Schyns & Oliva, 1994; for a review see Hegdé, 2008; Kauffmann et al., 2014), there is also considerable experimental evidence of a predominant processing of HSF information even for very short stimuli exposure duration (Campagne et al., 2016; Harel & Bentin, 2009; Morrison & Schyns, 2001; Rotshtein, Schofield, Funes, & Humphreys, 2010; Schyns, 1998; Schyns & Oliva, 1999). Furthermore, many studies have shown that the use of spatial frequency information during the processing of complex stimuli such as scenes is highly flexible and depends on many factors such as stimulus exposure duration (Schyns & Oliva, 1994), category (e.g., Awasthi, Sowman, Friedman, & Williams, 2013; Collin & McMullen, 2005; Rotshtein et al., 2010; Vannucci, Viggiano, & Argenti, 2001), or task constraints (Abrams, Barbot, & Carrasco, 2010; Campagne et al., 2016; Caplette, West, Gomot, Gosselin, & Wicker, 2014; Fradcourt, Peyrin, Baciú, & Campagne, 2013; Morrison & Schyns, 2001; Ozgen, Payne, Sowden, & Schyns, 2006; Schyns & Oliva, 1999; Sowden, Özgen, Schyns, & Daoutis, 2003). For example, different spatial frequency bands would be used according to their diagnosticity to categorize a specific visual stimulus. It has been shown that spatial frequencies below 2 cpd are diagnostic to perform basic-level categorization of scenes (e.g., forest, highway, mountain; McCotter, Gosselin, Sowden, & Schyns, 2005). However, intermediate spatial frequencies of 2.3–4 cpd would be required for basic-level categorization of objects (Caplette et al., 2014). Importantly, most of

these studies focused on the preferential use, the relevance, or the diagnosticity of a specific spatial frequency band for a given timescale or task constraint. However, how information from a spatial frequency band that is not directly relevant for the task at stake can contribute to categorization has received little interest. For example, in the context of a predominant coarse-to-fine sequence of spatial frequency processing, little is known about the relative role of HSF information and its impact on LSF processing. In their study, Mu and Li (2013) reported the results of an independent control experiment in which participants were instructed to attend and categorize the LSF scene in hybrid and ignore the HSF scene. In this experiment, they also observed a semantic interference effect suggesting that information from HSF also influenced LSF categorization. However, participants performance in the semantically dissimilar condition was not significantly different than chance, possibly indicating that they were not able to perform the task. Additionally, the authors did not address whether the semantic interference effect was greater in one attentional condition than the other (i.e., Attention to the LSF or HSF scene), that is, whether LSF interfered/influenced more strongly HSF categorization than the other way around. It is also likely that the relative weight of information from LSF and HSF varies over the time-course of scene categorization. In that sense, the influence of LSF information might be prominent at early stages of scene categorization, whereas information from HSF would take over at later stages.

The aim of the present behavioural study was to further examine how information from LSF and HSF interact and influence each other during visual scene categorization. To this end, we used a semantic interference paradigm (Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013). We presented hybrid stimuli made by combining the LSF and HSF components of two different scene categories, and we manipulated the semantic and physical similarity between the two scenes. Critically, we also manipulated attentional constraints. One group of participants had to attend and categorize the HSF scene in hybrids (HSF attentional condition), whereas the other group had to attend and categorize the LSF scene (LSF attentional condition). Additionally, to examine whether the relative weight of LSF and HSF varies over time, we manipulated the exposure duration of stimuli which were

presented for 30 or 100 ms. Prior to the attentional task, all participants had to categorize the individual LSF and HSF scenes used in hybrids in order to examine the processing of filtered scenes independently of attentional constraints. This also allowed us to ensure that filtered scenes composing the hybrids could be accurately categorized at both exposure durations. We expected the semantic interference effect to be modulated by the attentional task and the exposure duration of stimuli. In the context of a predominant coarse-to-fine processing of scenes, we hypothesized that at short exposure duration the semantic interference effect would be greater in the HSF than LSF Attentional condition, suggesting greater interference of LSF over HSF categorization than of HSF over LSF categorization. However, we expected this effect to be weaker at longer exposure duration, or perhaps reversed in such a way that HSF would interfere more with LSF categorization than the other way around.

Method

Participants

A total of 52 right-handed participants (15 male; Mean age \pm SD: 22 ± 3 years, range: 18–30 years) with normal or corrected-to-normal vision and no history of neurological disorders were included in this experiment. All participants gave their informed written consent before participating in the study in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Stimuli and procedure

The stimuli used in the present study were black and white photographs (256-level grey-scales, 256×256 pixels) taken among the scenes of the Labelme database (Oliva & Torralba, 2001) and subtended 6×6 degrees of visual angle. In order to select our stimuli, we first classified the 2687 scenes from this database into two distinct categories: natural (e.g., forests, coasts, open countryside) and man-made (e.g., highways, streets, buildings). We then assessed the physical similarity of scenes in the database using (1) the correlation between images based on their pixel intensity values (PI correlation) and (2) the correlation

between their amplitude spectra (AS correlation, i.e., correlation between pixel intensity values of amplitude spectra; see Kauffmann, Bourgin, et al., 2015, for a similar procedure). We considered two scenes as physically similar if their PI and AS correlation coefficients were superior to 0.6 (only 1.59% of images pairs matched both criteria). We considered two scenes as physically dissimilar if their PI correlation coefficient was inferior to 0.01 and their AS correlation coefficient was inferior to 0.6 (only 1.37% of image pairs matched both criteria). Based on these computations, we selected among the pairs of images with similar physical properties 40 pairs of scenes semantically similar (20 for each category; PI: 0.75 ± 0.08 ; AS: 0.67 ± 0.03) and 40 pairs of scenes semantically dissimilar (PI: 0.73 ± 0.06 ; AS: 0.66 ± 0.03), and among the pairs of images with dissimilar physical properties 40 pairs of scenes semantically similar (20 for each category; PI: 0.01 ± 0.00 ; AS: 0.43 ± 0.06) and 40 pairs of scenes semantically dissimilar (PI: 0.01 ± 0.00 ; AS: 0.44 ± 0.05). This resulted in a total of 160 scenes that were used as stimuli. With respect to the physical similarity conditions, scenes were selected to have similar PI and AS correlation values for the semantically similar and dissimilar pairs of images. Furthermore, mean energy was similarly distributed across spatial frequencies for the semantically similar and dissimilar image groups (see Figure 1(d)).

Each scene was filtered to preserve either LSF or HSF information, using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA). Filtered scenes were obtained by multiplying the Fourier transform of the original images with Gaussian filters. The standard deviation of Gaussian filters was a function of the spatial frequency cut-off for a standard attenuation of 3 dB. We removed spatial frequencies above 4 cycles per degree of visual angle (cpd, i.e., 24 cycles per image) in LSF scenes¹ and below 6 cpd (i.e., 36 cycles per image) in HSF scenes. The resulting filtered scenes were then normalized to obtain a mean luminance of 0.5, for pixel intensity values between 0 and 1 (i.e., mean luminance of 128 on a grey-level scale). It should be noted that, for some images, this luminance normalization procedure resulted in obtaining luminance values below 0 or above 1. These values were reset to 0 and 1, respectively, after conversion of values into integer luminance values ranging from 0 to 255 (i.e., conversion into 8-bit data type). However, this did not affect the mean

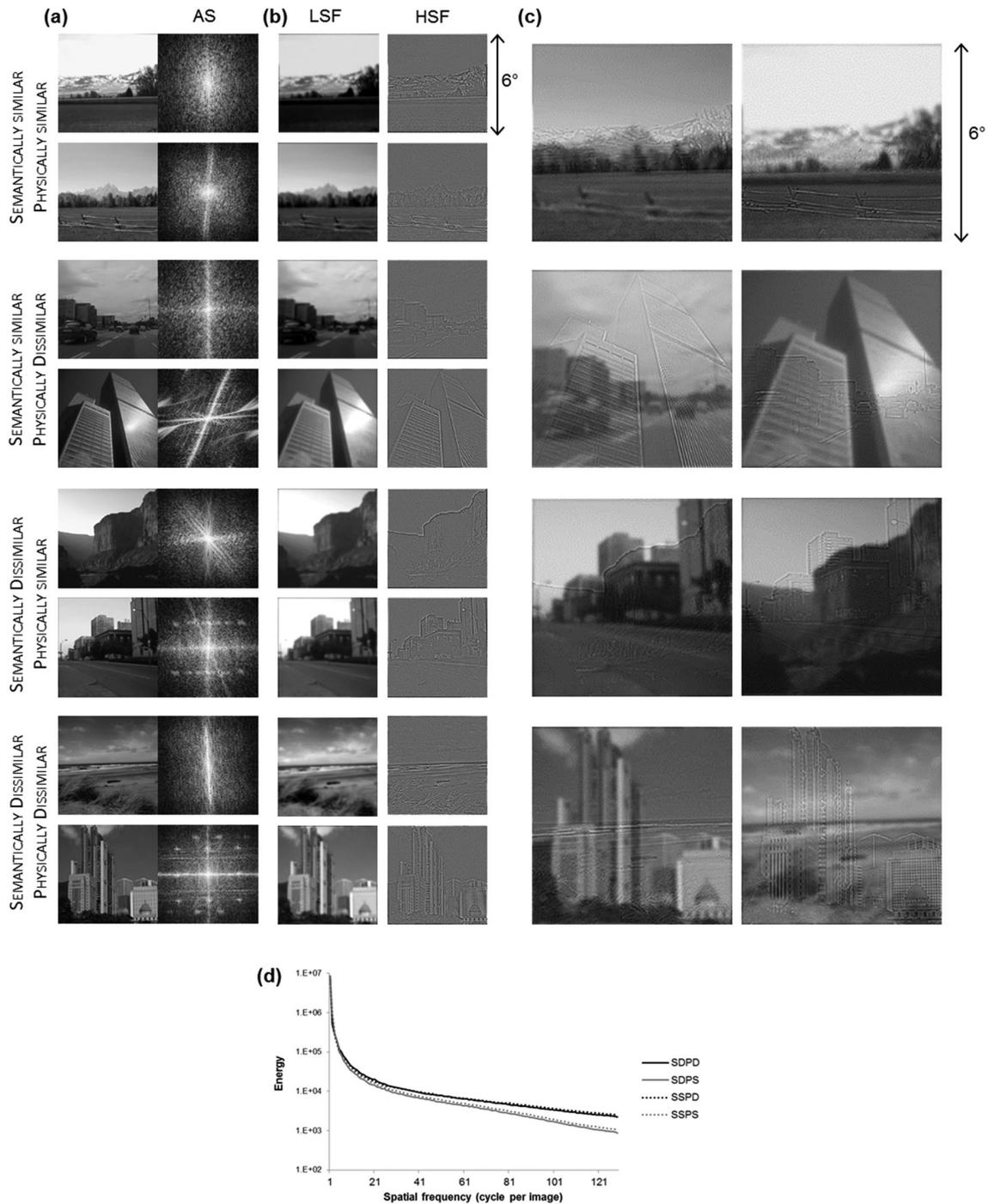


Figure 1. (a) Examples of black and white of photographs of scenes used with their amplitude spectrum (AS). On the amplitude spectrum images, low spatial frequencies (LSF) are close to the centre and high spatial frequencies (HSF) are on the periphery. (b) In the Individual scene experimental session (Session 1), we presented artificial and natural scenes filtered in LSF or in HSF. All participants had to categorize LSF and HSF scenes. (c) Examples of the stimuli used in the Hybrid experimental session (Session 2). Hybrid scenes were made by adding a LSF scene and a HSF scene. The two scenes in hybrids were either (1) semantically and physically similar (SSPS; e.g., LSF of a natural scene superimposed with HSF of another natural scene with similar physical properties), (2) semantically similar and physically dissimilar (SSPD; e.g., LSF of a man-made scene superimposed with HSF of another man-made scene with dissimilar physical properties), (3) semantically dissimilar and physically similar (SDPS; e.g., LSF of a natural scene superimposed with HSF of a man-made scene with similar physical properties), or (4) semantically and physically dissimilar (SDPD; e.g., LSF of a natural scene superimposed with HSF of a man-made scene with dissimilar physical properties). Half of the participants had to attend and categorize the LSF scene in hybrids (LSF-Attention) whereas the other half had to attend and categorize the HSF scene (HSF-Attention). (d) Mean energy distribution across spatial frequencies of images in the four experimental conditions (SDPD, SDPS, SSPD, SSPS).

luminance of filtered scenes which stayed at 0.5 (i.e., 128 on a grey-level scale) after this conversion. Root-mean square (RMS) contrast of filtered scenes was not modified (see note 1) and was higher for LSF than HSF scenes (LSF natural: mean $RMS \pm SD$: 0.24 ± 0.06 ; LSF man-made: 0.24 ± 0.05 ; HSF natural: 0.06 ± 0.02 ; HSF man-made: 0.07 ± 0.02). The displayed luminance of filtered scenes was slightly higher for LSF (mean luminance $\pm SD$: 14.89 ± 2.26 candela/m²) than HSF scenes (mean luminance $\pm SD$: 11.75 ± 1.33 candela/m²). RMS contrast of filtered scenes was higher for LSF than HSF scenes (mean $RMS \pm SD$: LSF: 13.29 ± 3.95 candela/m²; HSF: 3.23 ± 0.95 candela/m²). Hybrid stimuli were created by combining the LSF component in one scene with the HSF component in another scene. The two scenes which made up the hybrids could be either semantically similar (SS; e.g., LSF component of a natural scene superimposed with HSF component of another natural scene) or semantically dissimilar (SD; e.g., LSF component of a natural scene superimposed with HSF component of a man-made scene). Furthermore, the two scenes composing the hybrid could be either physically similar (PS) or physically dissimilar (PD). There were, therefore, four hybrid conditions (SSPS, SSPD, SDPS, and SDPD, see Figure 1) and 160 hybrid stimuli (40 per hybrid condition, see Supplemental data 1–4 for illustration of all stimuli). The 160 filtered scenes used to create the hybrids were also used in a separate experimental session where they were presented individually (either in LSF or HSF, see Figure 1(a)). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) on a 17-inch monitor, with a resolution of 1024 × 768 pixels, at a refresh rate of 80 Hz and with a viewing distance of 73 cm. In order to maintain the distance and the central position, participants' heads were supported by a chinrest.

All participants underwent two successive experimental sessions. In Session 1, they were presented with the individual LSF and HSF filtered scenes that were used to build up the hybrids. They were divided into two groups ($N=26$ each) for which scenes were presented for either 30 or 100 ms. LSF and HSF filtered scenes were randomly presented to the participants. Participants had to categorize the filtered scenes as natural or man-made. In Session 2, participants were presented with hybrid scenes. Each group of exposure duration was divided into two

groups ($N=13$ each) for which the attention was either directed on LSF or HSF. In the LSF Attentional group, they were instructed to ignore the HSF scene in hybrids, attend the LSF scene, and categorize it as natural or man-made. In the HSF Attentional group, they were instructed to ignore the LSF scene in hybrids, attend the HSF scene, and categorize it as natural or man-made. In total, there were four different groups according to the exposure duration of stimuli and the attentional task on hybrids (LSF-Attention/30 ms, LSF-Attention/100 ms, HSF-Attention/30 ms, and HSF-Attention/100 ms).

In both experimental sessions, each trial began with a central fixation point that was presented for 500 ms (in order to control the gaze direction to the centre of the screen) immediately followed by the stimulus (individual scene or hybrid) and a backward mask (built with 1/f noise) presented for 30 ms to prevent retinal persistence. All participants were requested to give their categorical answer as quickly and as accurately as possible, by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. Response keys were counterbalanced across participants. The experiment lasted about 20 min and response accuracy and reaction times (in ms) were recorded. Before the experiment, participants underwent a training session to get familiarized with tasks and stimuli.

Results

Data analysis

For both sessions we performed two ANOVAs on mean error rates (mER, in %) and mean correct response times (mRT, in ms). The ANOVAs in Session 1 included Attentional group in Session 2 (LSF-Attention, HSF-Attention) and Exposure duration (30 ms, 100 ms) as between-subject factors and Spatial frequency (LSF, HSF) of scenes as within-subject factor. It should be noted that the Attentional group factor was included in the analyses in order to assess whether there were any differences between the two attentional groups in Session 2 (LSF-Attention vs. HSF-Attention) during the processing of LSF and HSF scenes. For Session 2 the ANOVAs included Attentional condition (LSF-Attention, HSF-Attention) and Exposure duration of hybrids (30 ms, 100 ms) as between-subject factors and Semantic similarity

(semantically similar [SS], semantically dissimilar [SD]), and Physical similarity (physically similar [PS], physically dissimilar [PD]) of scenes in hybrids as within-subject factors. Because in some conditions error rates were close to ceiling, the ANOVA on mER for both sessions was performed after an Arcsine square root transformation, in order to reduce ceiling effects and ensure variance homogeneity. For both sessions, RT for each subject's correct response in each condition was trimmed in order to reduce the effect of extreme values, by removing responses inferior and superior to three times the interquartile interval. This resulted in removing an average 1.09% of the trials for Session 1 and 0.46% of the trials for Session 2. Analyses were performed using the Statistica 10.0 software (Statsoft, Tulsa, USA). Further pairwise comparisons were corrected using Bonferroni correction for the number of performed comparisons and corrected p -values are reported. Effect sizes were estimated by calculating the partial eta-squared (η^2). The significant level of tests was set at 0.05.

Session 1: LSF and HSF filtered scenes

The ANOVA performed on mER (see Figure 2(a)) revealed a main effect of Exposure duration ($F(1,48) = 54.24, p < .001, \eta^2 = 0.53$). Participants made more errors when scenes were presented for 30 ms (mean \pm SE: $16.30 \pm 8.69\%$) than during 100 ms ($4.47 \pm 4.39\%$). We also observed a main effect of Spatial Frequency ($F(1,48) = 11.04, p = .002, \eta^2 = 0.19$). Participants made more errors when categorizing HSF ($12.40 \pm 6.96\%$) than LSF scenes ($8.36 \pm 6.12\%$). Importantly, there was also a significant interaction between Exposure duration and Spatial frequency ($F(1,48) = 6.35, p < .05, \eta^2 = 0.12$). Participants made more errors when categorizing HSF than LSF scenes when presented for 30 ms (LSF: $12.69 \pm 8.18\%$; HSF: $19.90 \pm 9.20\%$, $p < .001$), but there was no difference in categorizing LSF and HSF scenes when presented for 100 ms (LSF: $4.04 \pm 4.07\%$; HSF: $4.90 \pm 4.72\%$, $p = 1$). There was no main effect of Attentional group ($F(1,48) < 1, p = .37$) and this factor did not interact with Exposure duration ($F(1,48) < 1, p = .40$), Spatial frequency ($F(1,48) < 1, p = .81$), or their interaction ($F(1,48) < 1, p = .86$). This ensures that differences observed between the categorization of LSF scenes in hybrids by the LSF-Attention group and the categorization of HSF scenes in hybrids in the HSF-Attention

group in Session 2 are not biased by differences between groups in spatial frequency processing in Session 1.

The ANOVA performed on mRT (see Figure 2(b)) revealed a main effect of Exposure duration ($F(1,48) = 19.56, p < .001, \eta^2 = 0.29$) and of Spatial Frequency ($F(1,48) = 25.41, p < .001, \eta^2 = 0.35$). Participants categorized the scenes more quickly when presented for 100 than 30 ms (30 ms: 670 ± 97 ms; 100 ms: 573 ± 65 ms) and when they were filtered in LSF than in HSF (LSF: 604 ± 72 ms; HSF: 639 ± 90 ms). There was no significant interaction between these two factors ($F(1,48) = 2.72, p = .11$). Furthermore, there was no main effect of Attentional group ($F(1,48) = 2.10, p = .15$), this factor did not interact with either Exposure duration ($F(1,48) = 2.72, p = .37$) or with Spatial frequency ($F(1,48) < 1, p = .93$), and the three-way interaction between these three factors was not significant ($F(1,48) = 1.68, p = .20$).

In summary, results of Session 1 showed that participants were able to accurately categorize the filtered scenes, even at short exposure duration. Results also indicated that participants performed better when categorizing LSF scenes than HSF scenes at short exposure duration of 30 ms and were overall faster in categorizing LSF than HSF scenes irrespective of their exposure duration. This suggests a LSF advantage (in terms of accuracy and reaction time) during scene categorization.

Session 2: hybrid scenes

The ANOVA performed on mER (see Figure 3(a)) revealed a main effect of the Semantic similarity between the scenes composing the hybrids ($F(1,48) = 61.28, p < .001, \eta^2 = 0.56$). Participants made more errors when the scenes in hybrids were semantically dissimilar ($36.13 \pm 13.84\%$) than similar ($19.76 \pm 15.02\%$). This indicated that when the two scenes were semantically dissimilar, the unattended scene was however processed at the semantic level and interfered with the categorization of the attended one (semantic interference effect). The main effect of Attentional condition was not significant ($F(1,48) < 1, p = .77$). However, this factor significantly interacted with the Semantic similarity of the two scenes composing the hybrids ($F(1,48) = 19.57, p < .001, \eta^2 = 0.29$). Pairwise comparisons revealed that participants made more errors when the two scenes in hybrids

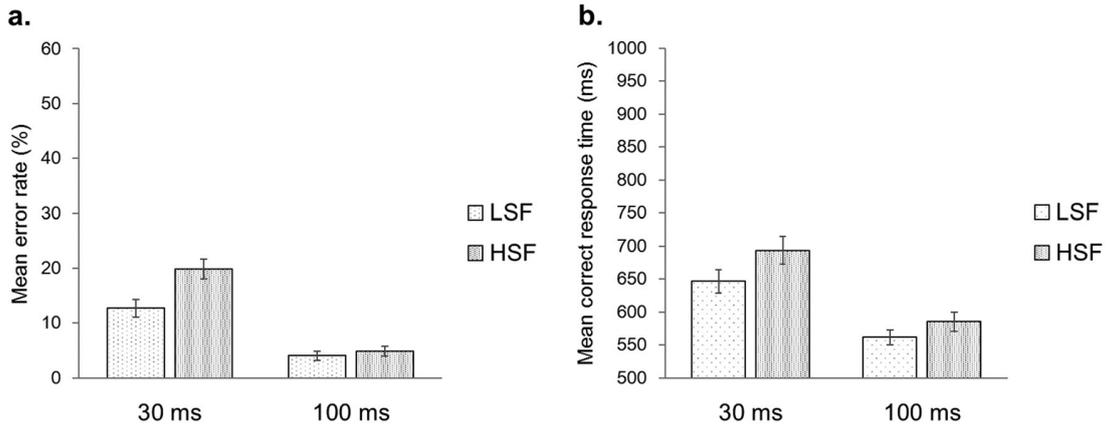


Figure 2. (a) Mean error rates (in %) and (b) mean correct response times (in ms) for the categorization of the individual scenes filtered in LSF and in HSF, for exposure duration of 30 and 100 ms. Error bars indicate standard error.

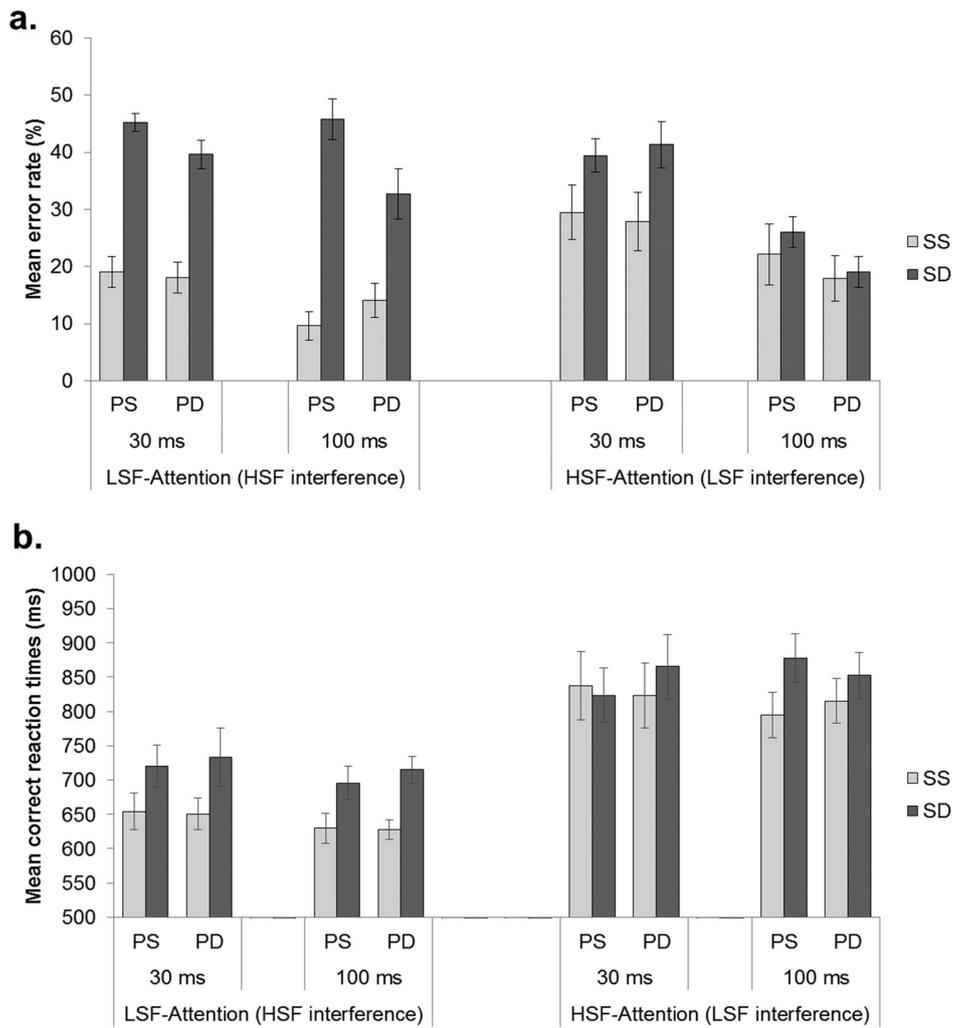


Figure 3. (a) Mean error rates (in %) and (b) mean correct reaction times (in ms) for the categorization of hybrids according to the Semantic similarity of scenes in hybrids (semantically similar-SS, semantically dissimilar-SD), their Physical similarity (physically similar-PS, physically dissimilar-PD), their Exposure duration (30 or 100 ms), and the Attentional condition (LSF-Attention, HSF-Attention). Error bars indicate standard error.

were semantically dissimilar than similar in both the LSF-Attention group (SS: $15.19 \pm 10.35\%$; SD: $40.82 \pm 11.37\%$; $p < .001$) and the HSF-Attention group (SS: $24.33 \pm 17.54\%$; SD: $31.44 \pm 14.31\%$; $p < .05$). Importantly, this semantic interference effect was greater in the LSF (25.63%) than HSF (7.12%) attentional condition group.

We also observed a significant main effect of Exposure duration of hybrid scenes ($F(1,48) = 12.98$, $p < .001$, $\eta^2 = 0.21$). Participants made more errors in categorizing hybrids when they were presented for 30 ms ($32.50 \pm 12.63\%$) than during 100 ms ($23.39 \pm 14.52\%$). However, the expected Attentional condition \times Exposure duration \times Semantic similarity interaction was not significant ($F(1,48) = 2.46$, $p = .12$).

Results analysis additionally revealed a significant interaction between the Physical similarity of scenes in hybrids, their Semantic similarity, and the Attentional condition ($F(1,48) = 9.77$, $p < .005$, $\eta^2 = 0.17$), due to a significant interaction between Physical and Semantic similarity in the LSF-attention condition ($p < .001$), but not in the HSF-attention condition ($p = 1$). Pairwise comparisons revealed that when participants had to attend the LSF scene in hybrids, the semantic interference effect was greater when the two scenes in hybrids were physically similar than dissimilar (SSPS = $14.33 \pm 10.45\%$; SDPS = $45.48 \pm 9.72\%$; $p < .001$; SSPD = $16.06 \pm 10.25\%$; SDPD = $36.15 \pm 13.02\%$; $p < .001$).

It should be noted that there was also a significant interaction between the Exposure duration of scenes in hybrids, their Semantic similarity, and their Physical similarity ($F(1,48) = 9.67$, $p < .05$, $\eta^2 = 0.17$). Pairwise comparisons showed that, for exposure duration of 100 ms, semantic interference effect was greater when the two scenes were physically similar than dissimilar ($p < .001$), whereas there was no interaction between semantic and physical similarity of hybrids for exposure duration of 30 ms ($p = 1$). Finally, the four-way Attentional group \times Exposure duration \times Semantic similarity \times Physical similarity interaction was not significant ($F(1,48) < 1$, $p = .36$).

The ANOVA performed on mRT (see Figure 3(b)) revealed a main effect of the Semantic similarity between the two scenes composing the hybrids ($F(1,48) = 56.65$, $p < .001$, $\eta^2 = 0.54$). Response times were slower when the two scenes composing hybrids were semantically dissimilar (785 ± 142 ms)

than similar (729 ± 147 ms), suggesting a semantic interference effect on response times. There was also a main effect of the Attentional condition ($F(1,48) = 24.10$, $p < .001$, $\eta^2 = 0.33$). Participants categorized the attended scene faster when it was in LSF (678 ± 93 ms) than in HSF (836 ± 143 ms) attentional condition. Importantly, there was a significant interaction between these two factors ($F(1,48) = 6.65$, $p < .05$, $\eta^2 = 0.12$). Pairwise comparisons revealed that participants categorized the scenes more slowly in the semantically dissimilar than similar condition, but this semantic interference effect was greater in the LSF-Attention condition (SS: 641 ± 78 ms; SD: 716 ± 108 ms, $p < .001$) than in the HSF-Attention condition (SS: 818 ± 147 ms; SD: 855 ± 140 ms, $p < .005$).

The expected Semantic similarity \times Exposure duration \times Attentional group was not significant ($F(1,48) = 2.11$, $p = .15$). However, we observed a significant four-way interaction between the Semantic similarity of scenes, their Physical similarity, their Exposure duration, and the Attention condition ($F(1,48) = 5.99$, $p < .05$). Pairwise comparisons revealed that when participants had to categorize the LSF scene in hybrids (LSF-Attention), scenes were categorized faster in the semantically similar than dissimilar condition for each Physical similarity condition (Physically similar and Physically dissimilar) of each Exposure duration condition (30 ms and 100 ms; all $ps < .05$). However, when participants had to categorize the HSF scene in hybrids (HSF-Attention), they categorized the scenes faster when they were semantically similar than dissimilar only for exposure duration of 100 ms and when the two scenes were physically similar (100 ms PSSH: 695 ± 119 ms; 100 ms PSSD: 877 ± 129 ms; $p < .001$; all other $ps > .05$).

Finally, to assess whether performances in categorizing LSF and HSF filtered scenes in Session 1 are linked to performances during categorization of hybrids (Session 2), we performed Pearson correlation analyses for mER and RTs between (1) categorization of LSF scenes in Session 1 and categorization of hybrids in the LSF-Attention condition in Session 2 and (2) categorization of HSF scenes in Session 1 and categorization of hybrids in the HSF-Attention condition in Session 2. Accuracy and reaction times for categorizing LSF scenes in Session 1 correlated positively with the ones for categorizing LSF scenes in hybrids in Session 2, irrespective of semantic similarity, physical similarity, and duration (mER: $r = 0.64$,

$p < .001$; mRTS: $r = 0.45$, $p = .021$). Similarly, accuracy and reaction times for categorizing HSF scenes in Session 1 correlated positively with the ones for categorizing HSF scenes in hybrids in Session 2, irrespective of semantic similarity, physical similarity, and duration (mER: $r = 0.53$, $p = .006$; mRTS: $r = 0.49$, $p = .011$).

Discussion

The present behavioural study aims at examining how information from low and high spatial frequencies interact and influence each other during scene categorization. To this end, we presented individual scenes and hybrids to participants who had to attend and categorize their LSF or HSF content. We manipulated the exposure duration of the stimuli which were presented for 30 or 100 ms. For hybrids stimuli, we also manipulated the semantic and physical similarity between the two scenes composing the hybrids. We used the semantic interference effect (i.e., poorer performance when the two scenes in hybrids are semantically dissimilar than similar) as the signature of the integration of semantic information from both spatial frequencies and as an estimate of their relative weight for scene categorization. Critically, this experimental paradigm allowed us to examine how information from a spatial frequency band that is not supposed to be explicitly processed implicitly influences processing of another spatial frequency content. Based on the hypothesis of a predominant coarse-to-fine processing sequence during scene categorization, we expected that information from LSF would weight more at short exposure duration (i.e., stronger interference of LSF over HSF scene categorization than the other way around), whereas information from HSF would take over with increasing exposure duration (i.e., stronger interference of HSF over LSF scene categorization than the other way around).

Low spatial frequency processing advantage

Results on individual LSF and HSF scenes (Session 1) revealed that participants were able to accurately categorize the LSF and HSF scenes (average error rate below 20%) when presented alone at all exposure durations. This allowed us to ensure that the semantic information contained in the LSF and HSF scenes

used in hybrids could be accurately extracted even at short exposure duration of 30 ms. Furthermore, we observed that LSF scenes were categorized more quickly than HSF scenes irrespective of the exposure duration of scenes, and more accurately than HSF scenes for the 30 ms condition only. These results are consistent with previous findings of a temporal precedence of LSF over HSF scene categorization observed in numerous studies using a simple categorization task of LSF and HSF stimuli (De Cesarei & Loftus, 2011; Kauffmann, Ramanoël, et al., 2015; Loftus & Harley, 2005; Parker et al., 1996). A similar pattern of results was found in the Hybrid experimental session (Session 2). Participants were also faster to categorize the attended scene when it was in LSF (LSF-Attention condition) than in HSF (HSF-Attention condition), irrespective of the semantic similarity of the unattended scene. Furthermore, performance for categorizing LSF and HSF individual scenes in Session 1 was positively correlated with the one for categorizing respectively LSF and HSF scenes in hybrids in Session 2. In other words, performances were consistent between Session 1 and Session 2 for the spatial frequency band that was explicitly processed and the LSF advantage observed in Session 1 remained in Session 2. Taken together, these results suggest that LSF information allows for accurate and rapid scene categorization. Furthermore, they support the view that, within a short amount of time, categorization is more efficient when based on LSF than HSF information (Schyns & Oliva, 1994), in accordance with a coarse-to-fine processing strategy. It is important to note that the cut-off frequency used to filter LSF scenes in the present study (4 cpd, i.e., 24 cpi) was relatively high as compared to the ones used in previous studies (usually around 2 cpd; see Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013; Schyns & Oliva, 1994). LSF scenes in the present study thus included a rather large part of the scene spatial frequency spectrum including low to intermediate spatial frequencies and therefore contained the most diagnostic features for scene and object categorization, which have been found to lie at 0–4 cpd (see Caplette et al., 2014; McCotter et al., 2005). Therefore, it is likely that the advantage for LSF processing observed in the present study also reflects the fact that the LSF scenes contained the most relevant information for categorization. It should also be noted that, in the present study, contrast was not equalized between

LSF and HSF filtered scenes. Because in natural scenes, luminance contrast typically decreases with increasing spatial frequencies (Field, 1987), LSF scenes were characterized by a higher contrast than HSF scenes. As it has been shown that reaction times decrease as contrast increases (Harwerth & Levi, 1978), it is therefore possible that the LSF advantage observed in the present study also results in part from a difference in contrast between LSF (mean displayed RMS contrast: 13.29 cd/m²) and HSF (mean displayed RMS contrast: 3.23 cd/m²) filtered scenes. The relative contribution of differences in contrast between spatial frequencies for a coarse-to-fine advantage during rapid scene categorization was the focus of a recent study conducted by our group (Kauffmann, Chauvin, et al., 2015). We used a design allowing us to examine separately the effect of spatial frequencies, contrast, and their combination, while participants performed a categorization task on sequences depicting a coarse-to-fine or a fine-to-coarse processing. Our results revealed an advantage for categorizing coarse-to-fine sequences relative to fine-to-coarse sequences. Importantly, we observed that, although this advantage was in part driven by differences in contrast between spatial frequencies, it predominantly relied on differences in terms of spatial frequencies.

Spatial frequency semantic interference

Results in the Hybrid experimental session (Session 2) additionally revealed that the categorization of scenes was impaired when the two scenes in hybrids were semantically dissimilar than similar. This result indicates that, although it was irrelevant to the task, the semantic information contained in the unattended scene in hybrids was processed and interfered with the categorization of the attended one. This semantic interference effect thus suggests the integration of semantic information contained in both scenes in hybrids. Furthermore, as expected, the semantic interference effect was modulated by the attentional constraints, i.e., by the spatial frequency content of the attended/unattended scenes in hybrids. Consistent with previous studies (Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013), participants exhibited a semantic interference of LSF information on the categorization of attended HSF scenes in hybrids (HSF-Attention condition). In an original way, the present study also

revealed a semantic interference of HSF information on the categorization of attended LSF scenes (LSF-Attention condition), and that the semantic interference effect was greater in the LSF than HSF attentional condition, suggesting that information from HSF interfered more strongly with LSF scene categorization than the other way around. These results indicate that, when there is a semantic conflict within the stimulus, information from HSF weights more than LSF information for achieving scene categorization. We did not find the expected modulation of the semantic interference effect by the exposure duration of stimuli. Results rather indicated that the semantic interference effect was always stronger in the LSF than HSF attentional condition, irrespective of the exposure duration of stimuli.

This last result thus partially contradicts our hypotheses. Indeed, we expected that LSF would weight more than HSF at short exposure duration, and that the weight of HSF would increase with exposure duration. Although our results were not consistent with these predictions, we however believe that they do not necessarily speak against a predominant coarse-to-fine processing strategy during scene categorization. First, numerous studies have previously shown that, although LSF information is preferentially used at short exposure duration to enable efficient and rapid scene categorization, it does not follow that it is always used first to support visual recognition. Indeed, and as suggested by our results, HSF can also be extracted very early in the visual processing time-course, and their preferential use would be determined by the demands of the visual task (Campagne et al., 2016; Harel & Bentin, 2009; Morrison & Schyns, 2001; Oliva & Schyns, 1997; Ozgen et al., 2006; Rotshstein et al., 2010; Schyns, 1998; Schyns & Oliva, 1994, 1999). For example, in Schyns and Oliva (1994), the perception of hybrids at short exposure duration of 30 ms was dominated by the HSF scene in 28% of the trials. Further studies from the same authors (Schyns & Oliva, 1999) showed that, during the categorization of hybrid faces presented for 50 ms, HSF were preferentially used to determine whether the face was expressive or not whereas LSF were preferentially used to determine the nature of the emotion. Overall, these results suggested that all spatial frequencies are available very early for categorization,

and that their selection may depend on interactions between the perceptual information available and the demands of a given visual task, thus speaking against a strictly coarse-to-fine processing sequence. Interestingly, it is very likely that the relatively greater interference of HSF information observed in the present study reflects such flexible use of spatial frequency information according to the task constraints – in our case the presence of an apparent conflict within the stimulus. Indeed, when there was no ambiguity within the stimulus (LSF and HSF scenes in Session 1), participants performed better when categorizing LSF than HSF scenes, especially at short exposure duration. As previously mentioned, this indicates that, by default, processing of LSF might be more efficient than that of HSF to rapidly access the scene category. However, when semantic information contained in LSF and HSF was incongruent (semantically dissimilar hybrids in Session 2), participants were interfered more strongly by the unattended scene when it was in HSF (LSF-Attention) than in LSF (HSF-Attention). As previously mentioned, the LSF scenes in hybrids had a higher contrast than the HSF scenes and included intermediate spatial frequencies diagnostic for categorization. The greater semantic interference in the LSF-Attention condition thus suggests a critical and irrepressible influence of HSF information for categorization, despite a priori more salient and relevant information contained in the LSF scene. Overall, results suggest that when there is uncertainty or ambiguity about the category of the visual stimulus, the processing of fine details contained in HSF might be the most relevant and reliable to disentangle between alternative interpretations and would thus strongly weight for scene categorization. It should be noted as a limitation of the present study that due to the hybrid methodology employed, potentially important spatial frequencies were not considered in the stimuli (i.e., spatial frequencies of 4–6 cpd). It is thus possible that these spatial frequencies also weight for scene categorization. Further studies explicitly manipulating the cut-off frequencies of filtered scenes in hybrids would allow us to provide further insight on that question and refine the present results. It is also noteworthy that overall, participants took longer to categorize the attended scene in hybrids when it was in HSF (HSF-Attention) than in LSF (LSF-Attention), irrespective of the semantic similarity of the

unattended scene. This suggests that fast categorization is more efficient when based on LSF and that more time is needed to process and categorize HSF. It is thus possible that the semantic interference of LSF over HSF categorization in the HSF-Attention condition might have been less evident due to initially poorer performances in that condition.

Effect of physical similarity

Our results additionally revealed that the semantic interference effect was modulated by the physical similarity between the two scenes composing the hybrids. In particular, the semantic interference effect was greater when the two scenes in hybrids were physically similar than dissimilar, especially for exposure durations of 100 ms. In their study, Mu and Li (2013) also reported such a modulation of the semantic interference effect by the physical similarity of scenes in hybrids, suggesting that the spatial overlap between LSF and HSF caused even greater impairment of scene categorization in the absence of any semantic congruence between the two scenes. Indeed, it is likely that, in the physically similar condition, the spatial overlap between the LSF and HSF scenes in hybrids might have hindered perceptual categorization performed on the basis of the attended scene physical properties. This could have resulted in enhanced semantic processing of the scenes in order to perform the task, especially when more time was available, resulting in even longer reaction times and error rates when the two scenes were semantically dissimilar. This would indicate that the processing of physical properties of LSF and HSF in scenes plays an important role and actively interacts with the processing of their semantic content over the time-course of scene categorization.

Impact for neurobiological models of scene perception

The semantic interference of LSF over HSF scene categorization was previously interpreted in the context of the proactive model of visual recognition proposed by Bar and colleagues (Bar, 2003, 2007; Bar et al., 2006; Kveraga, Boshyan, & Bar, 2007; Trapp & Bar, 2015). According to this model, fast processing of LSF information allows us to generate coarse predictions about the nature of the visual input in the

orbitofrontal cortex. These predictions would enable us to reduce the number of alternative interpretations of the stimulus and thereby tune the bottom-up processing of HSF via top-down influence to the inferotemporal cortex. Critically, this model thus assumes that it is the bottom-up processing of HSF information that is ultimately used to refine the coarse predictions and achieve recognition. Results of the present study, however, revealed that LSF information interfered less strongly with HSF categorization (HSF-Attention condition) than did HSF information with LSF categorization (LSF-Attention condition). In line with the above-mentioned interpretation, this may suggest that, in case of incongruence between LSF and HSF semantic information, it might be difficult and counterproductive to ignore the evidence accumulated in HSF against LSF-based coarse predictions. In that sense, the semantic interference of LSF over HSF processing would reflect the effect of erroneous top-down LSF-based predictions (Kauffmann, Bourgin, et al., 2015), whereas the semantic interference of HSF over LSF would rather reflect persistent and over-taking bottom-up integration of HSF information against these rapid LSF-based predictions.

Conclusions

To conclude, results of the present study support the current model of visual perception (Bar, 2003; Bullier, 2001; Kauffmann et al., 2014; Peyrin et al., 2010) suggesting that, during scene categorization, fast processing of coarse LSF information allows us to efficiently access the scene category and influence further processing of fine HSF information. Our results further allow us to refine them, by showing that semantic processing of HSF information can also occur very early and override processing of LSF information when there is a semantic conflict within the stimulus, thereby highlighting their critical role in ultimately mediating scene categorization. Overall, results of the present study indicate that the relative weight of LSF and HSF information during scene categorization varies according to the task constraints and the properties of the visual stimulus.

Note

1. We conducted a pilot study on 29 participants (1) to replicate the results of Mu and Li (2013) and Kauffmann,

Bourgin, et al. (2015) when participants attend and categorize the HSF scene in hybrids, but also (2) to consider the reverse attentional condition when participants attend and categorize the LSF scene in hybrids. We removed spatial frequencies content above 2 cpd (vs. 4 cpd in the present study). This value was chosen based on Schyns and Oliva's (1994) study on spatial frequency processing during scene categorization. Contrast was either not modified (LUM filtered scenes) or equalized between LSF and HSF to obtain a RMS contrast of 0.1, i.e., 25.6 on a grey-level scale (RMS filtered scenes) using the MATLAB image processing toolbox. Hybrid stimuli were then created by combining a LSF and a HSF scene for which the contrast was not modified (LUM hybrids) and by combining a LSF and a HSF scene for which the contrast was equalized (RMS hybrids). LUM and RMS hybrids were presented in distinct experimental blocks (counterbalanced between participants). Fourteen participants had to attend and categorize the HSF scene in hybrids and 15 participants had to attend and categorize the LSF scene in hybrids for both the LUM and RMS conditions. Results showed that the mean error rate of the SDPS condition of the LSF Attentional group was equal to $50\% \pm 10$ for LUM hybrids (not significantly different from the chance level; $t(14) = 0, p = 1$) and to $64\% \pm 20$ for RMS hybrids (significantly below chance; $t(14) = 2.77, p < .05$). These results suggest that participants were not able to categorize the LSF scene in hybrids and when the contrast of the HSF scene was increased in RMS hybrids. It should be noted that contrast normalization induces severe modifications in the amplitude spectrum properties of scenes that may bias behavioural responses. For example, the enhancement of contrast in HSF scenes using RMS normalization induces an additional unnatural LSF content that could directly mask the LSF scene of hybrids. Therefore, based on these pilot study's results, the present study was conducted without RMS contrast normalization between LSF and HSF scenes, and we chose to increase the spatial frequency cut-off of LSF scenes at 4 cpd. Using these values, participants had performances above chance in the SDPS condition of the LSF Attentional group ($45.48\% \pm 9.72, t(25) = -2.37, p < .05$). We also included a first experimental session to ensure that participants were able to categorize filtered scenes when they were not embedded in hybrids and independently of attentional constraints.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Agence Nationale de la Recherche grant (ANR-12-JHS2-0002-01 RECOR).

References

- Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception & Psychophysics*, 72(6), 1510–1521. doi:10.3758/APP.72.6.1510
- Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: Neuromagnetic and behavioural findings. *Frontiers in Human Neuroscience*, 7, 91. doi:10.3389/fnhum.2013.00091
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4), 600–609.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. doi:10.1016/j.tics.2007.05.005
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449–454. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1326160&tool=pmcentrez&rendertype=abstract>
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2-3), 96–107. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11690606>
- Campagne, A., Fradcourt, B., Pichat, C., Baciú, M., Kauffmann, L., & Peyrin, C. (2016). Cerebral correlates of emotional and action appraisals during visual processing of emotional scenes depending on spatial frequency: A pilot study. *Plos One*, 11(1), e0144393. doi:10.1371/journal.pone.0144393
- 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*. doi:10.3389/fpsyg.2014.00512
- Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics*, 67(2), 354–364. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15971697>
- De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic Bulletin & Review*, 18(5), 840–847. doi:10.3758/s13423-011-0133-6
- De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379–2394.
- Fradcourt, B., Peyrin, C., Baciú, M., & Campagne, A. (2013). Behavioral assessment of emotional and motivational appraisal during visual processing of emotional scenes depending on spatial frequencies. *Brain and Cognition*, 83(1), 104–113. doi:10.1016/j.bandc.2013.07.009
- 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
- Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision Research*, 18(11), 1579–1586. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24333686>
- Hegd e, 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
- 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
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of Cognitive Neuroscience*, 1–10. doi:10.1162/jocn
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49–57. doi:10.1016/j.visres.2014.11.013
- Kauffmann, L., Ramano el, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95. doi:10.1016/j.neuroimage.2015.02.058
- Kauffmann, L., Ramano el, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8(37), 1–14. doi:10.3389/fnint.2014.00037
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience*, 27(48), 13232–13240. doi:10.1523/JNEUROSCI.3481-07.2007
- Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, 12(1), 43–65. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15948283>
- Maunsell, J., Ghose, G. M., Assad, J. A., Mcadams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, 16, 1–14.
- McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, 12(6), 938–953. doi:10.1080/13506280444000599
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, 8(3), 454–469. doi:10.3758/BF03196180
- Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research*, 227(3), 367–377. doi:10.1007/s00221-013-3517-1
- 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
- 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, 72–107. doi:0010-0285/97
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.
- Ozgen, E., Payne, H. E., Sowden, P. T., & Schyns, P. G. (2006). Retinotopic sensitisation to spatial scale: Evidence for flexible spatial frequency processing in scene perception. *Vision Research*, 46(6-7), 1108–1119. doi:10.1016/j.visres.2005.07.015
- Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, 21(2), 147–160. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1513664>
- Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1448–1466. doi:10.1037/0096-1523.22.6.1448
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... Vuilleumier, P. (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
- Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, 11(5), 368–377. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4623891>
- Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision: It is not the stimuli but the comparison. *Journal of Vision*, 10(10), 25–25. doi:10.1167/10.10.25.Introduction
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, 67(1-2), 147–179. doi:10.1016/S0010-0277(98)00016-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.
- Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3), 243–265. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10193048>
- Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, 42(22), 2547–2554. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12445848>
- Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, 8(1978), 547–581.
- Sowden, P. T., Özgen, E., Schyns, P. G., & Daoutis, C. (2003). Expectancy effects on spatial frequency processing. *Vision Research*, 43(26), 2759–2772. doi:10.1016/S0042-6989(03)00480-2
- Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, 10, 538–545. doi:10.1016/j.tics.2006.10.007
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339(1), 190–198. doi:10.1111/nyas.12680
- Van Essen, D., & Deyoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). Cambridge: Bradford Book.
- Vannucci, M., Viggiano, M. P., & Argenti, F. (2001). Identification of spatially filtered stimuli as function of the semantic category. *Cognitive Brain Research*, 12(3), 475–478. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11689308>

Annexe 2

Influence of peripheral vision on object categorization in central vision

Alexia Roux-Sibilon

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Audrey Trouilloud

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Louise Kauffmann

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France
University Grenoble Alpes, CNRS, Grenoble INP,
GIPSA-lab, Grenoble, France

Nathalie Guyader

University Grenoble Alpes, CNRS, Grenoble INP,
GIPSA-lab, Grenoble, France

Martial Mermillod

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Carole Peyrin

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Predictive models of visual recognition state that predictions based on the rapid processing of low spatial frequencies (LSF) may guide the subsequent processing of high spatial frequencies (HSF). While the HSF signal necessarily comes from central vision, most of the LSF signal comes from peripheral vision. The present study aimed at understanding how LSF in peripheral vision may be used to generate predictive signals that guide visual processes in central vision. In two experiments, participants performed an object categorization task in central vision while a semantically congruent or incongruent scene background was displayed in peripheral vision. In Experiment 1, results showed a congruence effect when the peripheral scene was displayed before the object onset. In Experiment 2, results showed a congruence effect only when the peripheral scene was intact, thus carrying a semantic meaning, but not when it was phase-scrambled, thus carrying only low-level information. The study suggests that the low resolution of peripheral vision facilitates the

processing of foveated objects in the visual scene, in line with predictive models of visual recognition.

Introduction

Visual recognition in humans is considerably efficient and fast. Complex stimuli such as objects and natural scenes are robustly processed and categorized despite their infinite variability. Data on the functional neuroanatomy of visual pathways (Van Essen & DeYoe, 1995) and neurophysiological recordings in primates (De Valois, Albrecht, & Thorell, 1982; Shams & Von Der Malsburg, 2002; Shapley & Lennie, 1985) suggest that the visual system is able to rapidly extract the low spatial frequencies (LSF) from a visual scene through fast magnocellular pathways. LSF signals convey coarse information about the global shape and

Citation: Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, 19(14):7, 1–16, <https://doi.org/10.1167/19.14.7>.

<https://doi.org/10.1167/19.14.7>

Received May 22, 2019; published December 11, 2019

ISSN 1534-7362 Copyright 2019 The Authors



structure of the scene. The processing of this information precedes the processing of a high spatial frequency (HSF) signal, which conveys finer information about the scene such as edges and object details through slowest parvocellular pathways. Accordingly, theories of visual perception have proposed that visual information is integrated in a *coarse-to-fine* manner (Bar, 2003; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Schyns & Oliva, 1994).

In this theoretical framework, it was also hypothesized that the brain uses the rapidly available LSF to predict visual inputs, especially within the orbitofrontal cortex (Bar & Aminoff, 2003; Kauffmann et al., 2014; Kveraga, Boshyan, & Bar, 2007; Peyrin et al., 2010). The predictive signal would be back-projected, via top-down connections, to occipito-temporal visual areas to guide bottom-up processes. Predictions could then influence the subsequent processing of HSF. As experimental evidence, a combined magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) study (Bar et al., 2006) demonstrated earlier activations in the orbitofrontal cortex than in the occipito-temporal cortex during recognition of object images, this early activity depending on the presence of LSF in the image. Recent fMRI studies investigating the effective connectivity between these regions using dynamic causal modeling showed that a magnocellular signal (e.g., achromatic and low-luminance contrast drawings, LSF-filtered scenes) increases the connectivity strength from the orbitofrontal cortex to the inferotemporal cortex (Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007). Petras, ten Oever, Jacobs, and Goffaux (2019) used a classifier trained to discriminate between EEG scalp patterns evoked by LSF inputs and EEG scalp patterns evoked by HSF inputs in order to tease apart LSF and HSF contribution to the neural response evoked by broadband stimuli. They then tested the classifier on EEG scalp patterns evoked by intact and phase-scrambled broadband faces. Results showed an early LSF dominance followed by a reduced HSF dominance in response to intact, but not to scrambled, face stimuli. Given the fact that LSF was informative about HSF only in intact images, the reduction in the late contribution of HSF suggests that the first LSF parsing elicits a robust representation of the face that subsequently reduces (unnecessary) HSF processing.

However, these studies often use small stimuli displayed in central vision and therefore miss the selectivity of peripheral vision for LSF processing. Indeed, the processing of spatial frequencies is nonhomogeneous throughout the visual field (Curcio & Allen, 1990; Curcio, Sloan, Kalina, & Hendrickson, 1990). In the retina, the density of midget ganglion cells, tuned to HSF, is greater in the fovea, while the density of parasol ganglion cells, tuned to LSF, increases with

retinal eccentricity. Therefore, the receptive fields in peripheral vision are too large to capture HSF signals. The counterpart is that the visual system has a considerable amount of LSF information rapidly available in peripheral vision that can contribute to the generation of predictive signals. In this view, the processing of details in central vision could benefit from predictive signals originating from both the overlapping central LSF signal and the eccentric LSF signal.

The objective of the present study is to test whether LSF information in peripheral vision can be used to generate predictive signals that guide visual processes in central vision. In real-life conditions, observers often foveate relevant elements (e.g., a face during a social interaction, an object while performing an action) while these are embedded within a coherent context encompassing the whole visual field. Central vision is thus more suited for object perception and peripheral vision for scene perception. This distinction is found in the visual cortex where object-selective areas (in the lateral occipito-temporal cortex) respond more strongly to central rather than peripheral visual input, while scene-selective areas (in the medial occipito-temporal cortex) respond more strongly to peripheral rather than central visual input (Arcaro, McMains, Singer, & Kastner, 2009; Baldassano, Fei-Fei, & Beck, 2016; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Malach, Levy, & Hasson, 2002). In the present study, we asked participants to categorize an object in central vision while a semantically related (predictive/congruent) or unrelated (nonpredictive/incongruent) scene background was presented in peripheral vision. We did not filter out the spatial frequency content from the scene, since peripheral vision acts as a natural low-pass filter. We expected that LSF information available in the broadband scene displayed in peripheral vision would influence the explicit categorization of the object in central vision (resulting in better performance in congruent than in incongruent trials). A large number of studies investigated the interactions between objects and context (Bar & Ullman, 1993; Boucart, Moroni, Szaffarczyk, & Tran, 2013; Davenport, 2007; Davenport & Potter, 2004; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Rémy et al., 2013; Sun, Simon-Dack, Gordon, & Teder, 2011). Typically, categorization performance is better when objects are within a congruent rather than incongruent context. A congruent context can also facilitate object processing in patients with age-related macular degeneration, a retinal disease causing central vision loss due to the destruction of macular photoreceptors (Boucart et al., 2013). However, because the distinction between central and peripheral vision was not the aim of these studies, authors did not control either the position of the object on the background scene, or the retinal size

of the object and the scene (resulting in rather small scenes and big objects).

Here, we examined how scene information influences object categorization when object information is restricted to central vision and scene information to peripheral vision. We manipulated the time during which the scene was processed before object onset (Experiment 1). We expected that if information in peripheral vision is used to generate a predictive signal that shapes visual processing in central vision, the peripheral influence on object categorization (i.e., the congruence effect) would be greater when peripheral information can be accumulated and result in a sharper representation of the scene—that is, when the scene is presented longer before object onset. To test whether low-level aspects of scene processing are sufficient to influence visual processing in central vision, or whether achieving a semantic representation of the scene is needed (Bar, 2003; Kauffmann, Bourgin, Guyader, & Peyrin, 2015), we presented the scenes in two versions: intact and phase-scrambled (Experiment 2). The phase-scrambled version of the scene is made by disrupting the phase of the scene image while preserving low-level (spatial frequency and orientation distribution) information. We expected that if the semantic representation of the scene is used to generate predictions, the peripheral scene would influence object categorization when it is intact more than when its phase is scrambled.

Experiment 1

Methods

Participants

Fifty right-handed participants (43 women; $M \pm SD = 21 \pm 4$ years, range: 18–26 years) with normal or corrected-to-normal vision participated in the experiment. Most of them were psychology students who received course credits for their participation. They gave their informed written consent before participating in the study, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and approved by the local ethics committee (Grenoble-Alps Research Ethics Committee, Community University Grenoble Alps, IRB00010290).

Stimuli

Stimuli were images consisting of a combination of an object and a scene background. All images used to create the stimuli (scenes and objects) were downloaded from the image bank Pixabay (<https://pixabay.com/>) under Creative Commons Zero license, or from Google

Images searches of copyright-free pictures. The scenes used as backgrounds were photographs representing either outdoor natural landscapes (e.g., savannah, field, park ice), or indoor views (e.g., living room, corridor). They were converted into 256-level grayscale and rescaled to 1024×768 pixels. Stimuli were displayed on a 30-in. monitor (Dell Ultrasharp) with a resolution of 2560×1600 pixels and a refresh rate of 60 Hz. We estimated the gamma function of the monitor by measuring the luminance values of the display for different values of uniform gray-level stimuli generated in MATLAB (MathWorks, Natick, MA) with a calibration tool (Spyder5ELITE, Datacolor, Rotkreuz, Switzerland). Based on the estimated function, we gamma-corrected luminance values of each scene to linearize indirectly the gamma function. Mean luminance and root mean square (RMS) contrast (standard deviation of the luminance) of the scenes were then equalized to obtain a mean luminance of 0.51 for luminance values ranging from 0 to 1 (i.e., mean luminance of 130 on a gray-level scale) and a mean RMS of 0.24 (i.e., 61 on a gray-level scale). These values correspond to the average values of luminance and RMS contrast of all scene images (landscapes and indoor views). Objects were photographs of animals (e.g., cow, penguin) and pieces of furniture (e.g., armchair, coffee table) that were cut out of natural images using Adobe Photoshop. Each object was rescaled to 128 pixels on its larger side (either width or height, depending on the object).

We created 20 groups of four object-scene stimuli (80 stimuli) by combining one scene background with one object. The four combinations resulted in two congruent (animal + outdoor context and furniture + indoor context) and two incongruent (animal + indoor context and furniture + outdoor context) object-scene associations (Figure 1). Scenes and objects forming congruent associations were combined based on their real-world congruency (e.g., cow + field, penguin + park ice). Each object was pasted on the corresponding congruent and incongruent backgrounds, at the center on the horizontal axis, but in the lower part on the vertical axis (centered at 576 pixels from the top of the image) in order to respect the usual position of an object in a photograph. The object was isolated from the background scene by a circular gray patch (diameter of about 250 pixels) whose edges were smoothly blended into the scene. This allowed us to restrict contextual information to peripheral vision only, and to control for possible local contrast effects on the perception of the object. Stimuli can be download from <https://osf.io/mfhx5/>.

Procedure

The experiment was programmed on E-Prime 2.0 software (E-Prime Psychology Software Tools Inc.,

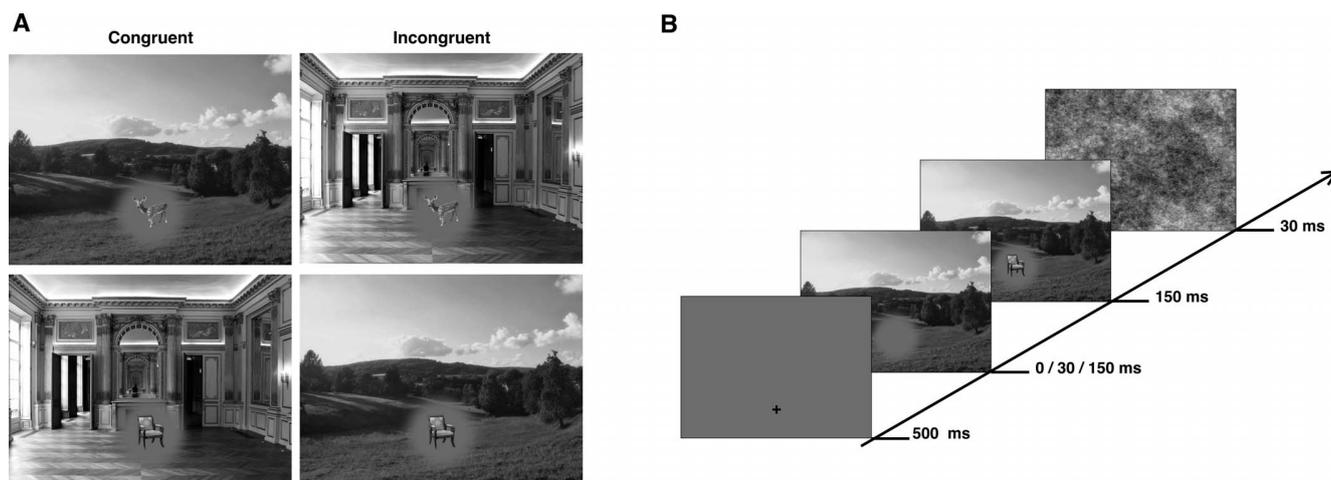


Figure 1. (A) Example of a group of four scene + object stimuli used in Experiment 1. (B) Schematic of the experimental procedure. In the SOA-0 condition, the scene + object stimulus appeared directly, without any pre-exposition of the peripheral scene.

Pittsburgh, PA). Participants viewed stimuli at a distance of 50 cm. At this distance, the background scenes subtended 24° of visual angle horizontally and 18° vertically. The larger side of the objects (either the width or the height) subtended 3° and the gray circular patch subtended about 6° in diameter and was centered at $\sim 13^\circ$ from the top of the image. The participant's head was supported by a chinrest in order to maintain distance.

Participants were instructed to maintain their fixation on a little white fixation cross, centered on the screen location at which the object was presented (i.e., at the center of the screen horizontally and in the lower part vertically). Each trial began with the fixation cross presented for 500 ms, followed by a stimulus sequence, and by a backward mask (1/f noise) for 30 ms. Interstimulus interval was of 2,500 ms. The stimulus sequence depended on the experimental condition. In the stimulus onset asynchrony (SOA)-0 condition, the scene-object stimulus was presented for 150 ms. In the SOA-30 condition, the scene-object stimulus was also presented for 150 ms, but was preceded by the scene image alone (including a circular gray patch masking central visual information) for 30 ms. Similarly, in the SOA-150 condition, the scene-object stimulus was presented for 150 ms, and was preceded by the scene image alone for 150 ms. The three scene SOA conditions were blocked and block order was counterbalanced across participants by a Latin square procedure. Participants had to categorize the object as animal or furniture by pressing two keys on a keyboard. They received no special instruction about the background, but if they questioned it, they were told that the background did not matter. Response keys were

counterbalanced across participants. Accuracy and response time (RT) were recorded at each trial. There were 240 trials (2 Congruence \times 3 Scene-SOA \times 20 Stimuli Groups \times 2 Object Categories). Participants could pause between scene SOA blocks, and the experiment lasted about 20 min. Prior to completing the experimental trials, participants underwent a short training session (12 trials) to be familiarized with the stimuli and the task.

Results

For each participant, correct RTs (in ms) were log transformed and then trimmed by removing trials for which RT was inferior or superior to the condition average ± 2.5 SDs (1.76% of trials were excluded). Mean correct RTs (mRT), mean log transformed correct RTs (mLog[RT]) and mean correct response rates (mCR), with standard deviations, for each experimental condition are reported in Table 1. We conducted two repeated measures ANOVA on mCR and mLog[RT] with congruence (congruent or incongruent stimuli) and scene SOA (SOA-0, SOA-30, SOA-150) as within-subject factors. Further pairwise comparisons were tested with two-tailed paired-samples t tests, and trend analyses with polynomial contrasts, using a Bonferroni adjustment of alpha level to correct for multiple tests ($0.05/8 = 0.006$ for the eight pairwise tests and trend analyses performed). Effect size was estimated by calculating Cohen's d for within subject design. Statistical analyses were conducted in R.

Analysis of mCR showed that mCR was larger in the congruent condition ($M \pm SD$: 0.96 ± 0.05) than in the incongruent condition (0.95 ± 0.05 ; $F[1, 49] = 6.52$, $p =$

	Congruent SOA-0	Incongruent SOA-0	Congruent SOA-30	Incongruent SOA-30	Congruent SOA-150	Incongruent SOA-150
mCR	0.96 ± 0.20	0.96 ± 0.20	0.97 ± 0.18	0.96 ± 0.20	0.96 ± 0.19	0.95 ± 0.23
mRT	567 ± 159	568 ± 158	539 ± 149	543 ± 140	527 ± 135	547 ± 130
mLog(RT)	2.74 ± 0.10	2.74 ± 0.10	2.72 ± 0.10	2.72 ± 0.10	2.71 ± 0.10	2.73 ± 0.09

Table 1. Mean correct response rate (mCR), mean correct response times in milliseconds (mRT), and mean log transformed correct response times (mLog[RT]), with standard deviations, for each experimental condition (congruence, scene SOA).

0.014, $d = 0.36$). The main effect of Scene-SOA was not significant ($F[2, 98] = 1.37$, $p = 0.258$), and this factor did not interact with the congruence of stimuli ($F[2, 98] = 2.22$, $p = 0.114$; Figure 2). It should be noted that the global mCR was high (0.96 ± 0.20), with distributions of individual observations reflecting a ceiling effect.

Analysis of mLog(RT) showed that participants were faster to categorize objects in congruent stimuli (2.723 ± 0.102) than in incongruent stimuli (2.731 ± 0.097 ; $F[1, 49] = 39.44$, $p < 0.001$, $d = 0.88$). The main effect of scene SOA was also significant ($F[1, 98] = 14.40$, $p < 0.001$, $d = 0.54$). Both linear and quadratic trends were significant (linear contrast: $t[49] = 4.22$, $p < 0.001$, $d = 0.60$; quadratic contrast: $t[49] = 2.83$, $p = 0.007$, $d = 0.40$), indicating that participants were faster to categorize the objects in both the SOA-30 (2.721 ± 0.010) and SOA-150 (2.719 ± 0.095) than in the SOA-0 condition (2.741 ± 0.103). Importantly, scene SOA interacted with the congruence of the stimuli ($F[1, 98] = 12.4$, $p < 0.001$, $d = 0.50$; Figure 2). We tested the difference between the congruent and incongruent trials for each scene SOA condition. It was significant for the SOA-150 condition (congruent: 2.710 ± 0.096 ; incongruent: 2.728 ± 0.092 ; $t[49] = 7.60$, $p < 0.001$, $d = 1.08$), but neither for the SOA-30 condition (congruent: 2.718 ± 0.104 ; incongruent: 2.723 ± 0.096 ; $t[49] = 2.68$, $p = 0.010$) nor for the SOA-0 condition (congruent: 2.740 ± 0.103 ; incongruent: 2.741 ± 0.102 ; $t[49] = 0.64$, $p =$

0.525). Post hoc trend analyses were conducted to examine the effect of scene SOA on congruent and incongruent stimuli separately. For the congruent condition, the linear trend was significant ($t[49] = 5.44$, $p < 0.001$, $d = 0.77$), but the quadratic trend was not significant ($t[49] = 1.89$, $p = 0.064$). This suggests that RT decreases linearly when scene SOA increases in the congruent condition. In the incongruent condition, the linear trend was not significant ($t[49] = 2.42$, $p = 0.019$) but the quadratic trend was ($t[49] = 3.36$, $p < 0.006$). We then tested for a linear modulation of the difference between incongruent and congruent trials by the scene SOA factor. The linear contrast was significant ($t[49] = 4.70$, $p < 0.001$, $d = 0.31$) and the quadratic contrast was not significant ($t[49] = 1.24$, $p = 0.223$), suggesting that the peripheral scene influence on object categorization increased linearly with longer scene SOA.

Results of Experiment 1 can be summarized following two main observations. First, contrary to our hypothesis, no effect of congruence was observed for the SOA-0 and SOA-30 conditions, suggesting that the representation of the scene was not robust enough to generate strong predictions in those conditions and thus influence the object categorization. Second, there was a congruence effect in the SOA-150 condition (on RTs). This effect suggests that when the scene background was sufficiently processed before the object onset, participants automatically processed information in peripheral

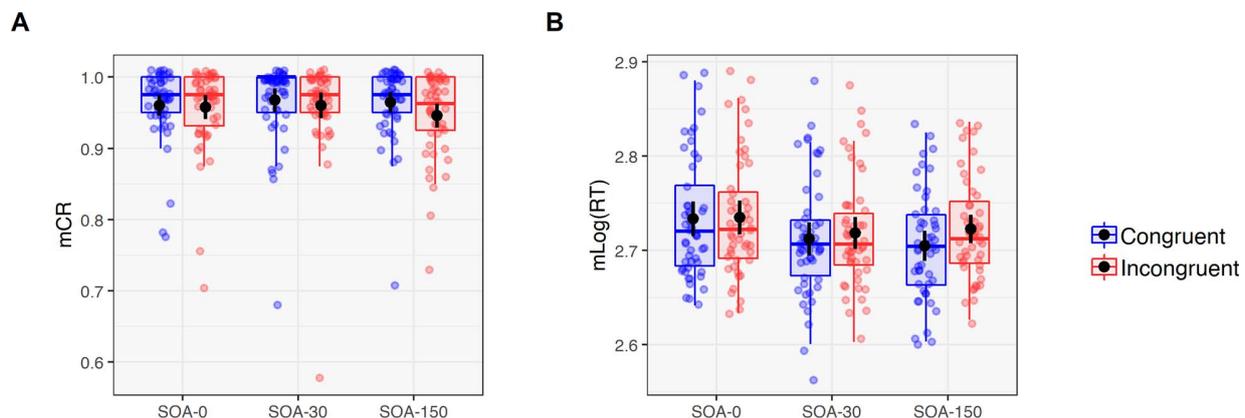


Figure 2. (A) Mean correct response rates (mCR) and (B) mean log transformed correct RTs (mLog[RT]) for the categorization of the object according to the congruence between the object and the scene in peripheral vision (congruent, incongruent) and the scene SOA (SOA-0, SOA-30, SOA-150). Black dots and error bars indicate means and 95% CIs. Color dots are individual observations (slightly jittered for better visualization).

vision although it was irrelevant to the task. This result is consistent with the hypothesis of predictive influences from peripheral vision. The effect, however, is rather small in practice, since the raw observed difference between the congruent and incongruent conditions was 20 ms. However, the effect seems to be robust and should be detected easily in subsequent experiments, as suggested by the large Cohen's d (1.08). Thus, results of Experiment 1 are consistent with the hypothesis of predictive influences from peripheral vision, but suggest that the influence is rather small. Moreover, the mechanisms underlying the congruence effect are not clear. First, as we did not test an experimental condition without a context, we cannot know whether object categorization was facilitated in the congruent condition or impaired in the incongruent condition. In addition, it is possible—at least in a certain proportion of trials—that the recognition of the scene category itself may have directly driven the response, without any feedback modulation on the processing of the object. An extreme example of such an “overhasty” response strategy would be a situation where a participant always responds *animal* when an outdoor scene is recognized and *furniture* when an indoor scene is recognized, regardless of the object. In this case, the proportion of correct responses would be 100% in the congruent condition and 0% in the incongruent condition, resulting in a maximum congruence effect. Considering this, the congruence effect measured in our experiment might not be entirely due to the *integration* of the peripheral visual input during the categorization of the object in central vision. Unfortunately, Experiment 1 did not allow us to dissociate an actual object categorization influenced by the scene from mere scene-consistent responses. Finally, we observed a ceiling effect of accuracy, suggesting that the design of this experiment made the task too easy, making it difficult to detect interactions on accuracy.

Experiment 2

We conducted a second experiment to understand better how peripheral vision influences central vision processing considering the limits described above. We used the same experimental paradigm as in Experiment 1, but the scenes were always presented 150 ms before the object onset (i.e., the condition for which we observed a significant congruence effect in Experiment 1). First, in order to rule out the possibility of a ceiling effect, we parametrically decreased the visibility of objects by manipulating a phase-scrambling parameter. Objects were presented at seven different levels of visibility. The parametric manipulation allowed us to map psychometric functions of mCR as a function of visibility levels. Another aim of this manipulation was

to promote the congruence effect. The rationale was that when the input is poor or ambiguous, the weight of predictions should be stronger compared to that of the inputs (Kok & de Lange, 2015). Therefore, information in peripheral vision should have a stronger influence when the central visual information is poor. In this context, studies have shown that predictive processes enhance visual perception when the visual stimulation is noisy or incomplete (e.g., Brandman & Peelen, 2017; Tang et al., 2018; Teufel, Dakin, & Fletcher, 2018; Wyatte, Curran, & O'Reilly, 2012). Secondly, in order to consider the direction of the congruence effect (facilitation in the congruent condition, hindrance in the incongruent condition, or both), we included a baseline condition. In this condition, we combined the object with a meaningless image background (1/f noise). Thirdly, in order to distinguish the part of the response based on the peripheral scene alone from the part of the response actually due to the influence of the peripheral scene on object categorization, we included trials for which no object was presented. In those trials, the object region was simply filled with 1/f noise. In that case, there was no correct response. However, since these trials were randomly embedded among object trials (some of them being of low visibility), participants did not realize the absence of objects. We tested if the tendency to rely on the scene when no object was present correlated with the congruence effect. A positive correlation would suggest that the congruence effect is partly due to a mere processing of the scene, without exerting feedback modulation during the processing of the object. In Experiment 2, we were also interested by the nature of the peripheral influence. In the peripheral visual field, the visual system extracts low-level visual features (spatial frequencies and orientations), whose processing allows the construction of high-order semantic representations. As can be seen in Figure 3, pieces of furniture and indoor scenes tend to have similar amplitude spectra. The energy is mainly distributed on vertical and horizontal, but also few oblique, orientations (due to the viewpoint perspective), ranging from the lowest to the highest spatial frequencies. In the same way, animals and outdoor scenes tend to have similar amplitude spectra, with energy more sparsely distributed throughout orientations, and mostly on the lowest spatial frequency range. It is thus possible that the influence of peripheral vision on central object categorization is only based on low-level visual features, rather than on a higher order semantic representation of the scene. Both influences are plausible. For example, predictive coding theories of vision (Friston & Stephan, 2007; Lee & Mumford, 2003; Rao & Ballard, 1999) propose that predictions flow between hierarchical areas within the visual cortex, where low-level aspects are represented. In predictive models of visual recognition (Bar, 2003; Kauffmann et

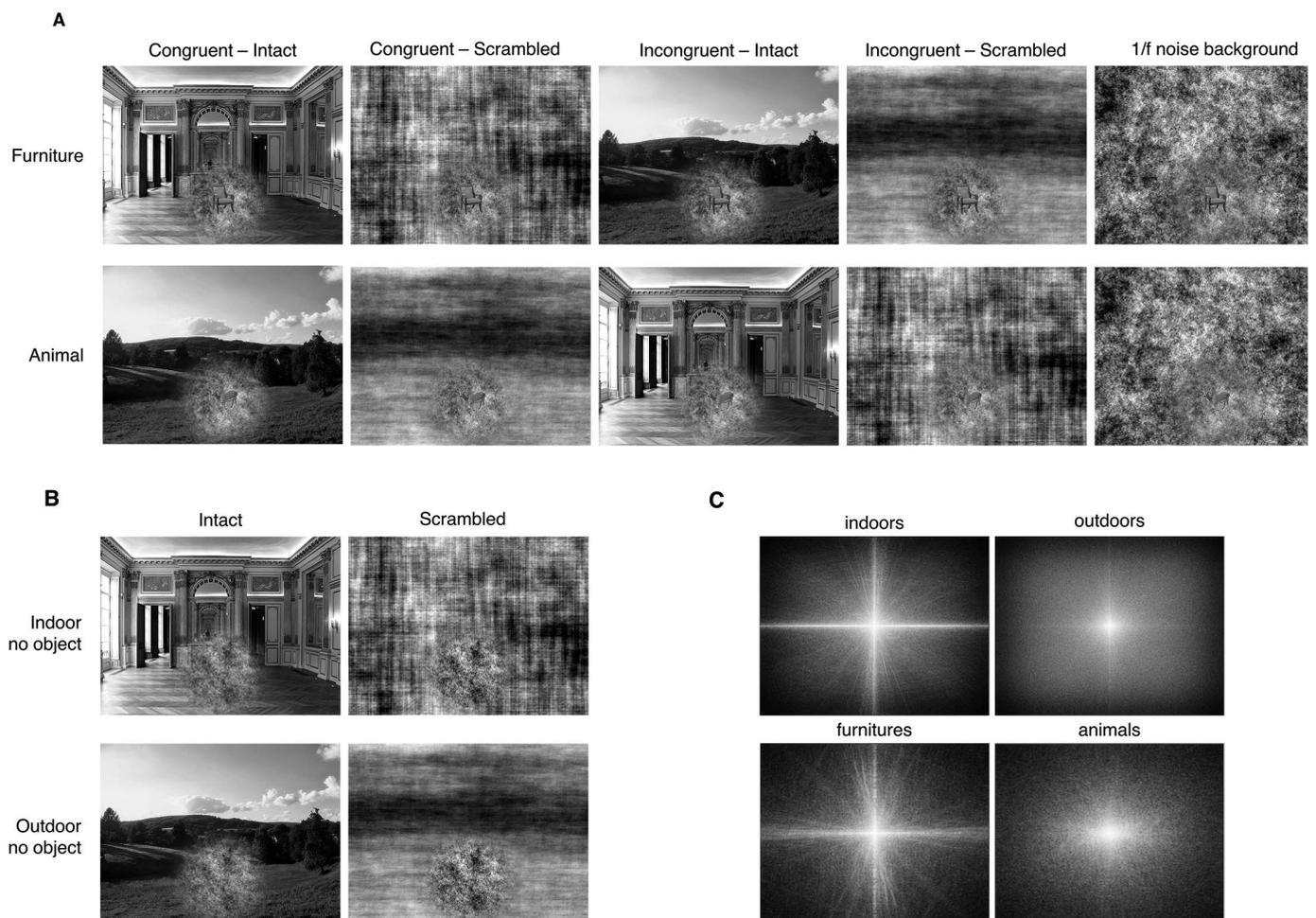


Figure 3. Example of a set of stimuli used in Experiment 2. (A) The object was a piece of furniture (first row) or an animal (second row) pasted on a 1/f noise and embedded in a congruent or incongruent scene background whose phase was either intact or scrambled, or in a meaningless 1/f noise background. (B) “Scene-alone” condition, in which no object was present in the 1/f noise. (C) The mean amplitude spectrum of indoor scenes is similar to that of furniture, while the mean amplitude spectrum of outdoor scenes is similar to that of animals. For illustration purposes, contrast and phase coherence of objects were slightly increased in the present figure. See Figure 4 for a zoom in on the object.

al., 2014; Kveraga et al., 2007; Peyrin et al., 2010), predictions are triggered in the inferofrontal cortex, where semantic aspects would be represented. To test for the influence of the two types of information (low-level and semantic) we manipulated the presence of a semantic content in the background scene. In the *intact* condition, we used the original scene image. In the *scrambled* condition, we suppressed the semantic information by scrambling the phase spectrum of the intact scene image via random permutation. This procedure is known to preserve orientation and spatial frequency content while preventing the processing of any semantic content (Goffaux et al., 2010; Woodhead, Wise, Sereno, & Leech, 2011). If the peripheral influence is due to low-level visual features only, we expected to observe an effect of congruence both in the intact and scrambled condition. On the contrary, if the

peripheral influence involves semantic representations as well, we expected to observe a greater effect of congruence in the intact condition than in the scrambled condition.

Methods

Participants

Eighteen right-handed participants (14 women; $M \pm SD$: 21 ± 3 years; range: 19–30 years) who did not participated in Experiment 1 were included. They were psychology students who received course credits for their participation. The sample size was chosen based on a power analysis with estimated effect size of 1.08 (effect size of the congruence effect in the SOA-150 condition in Experiment 1) to achieve power of 0.99 at

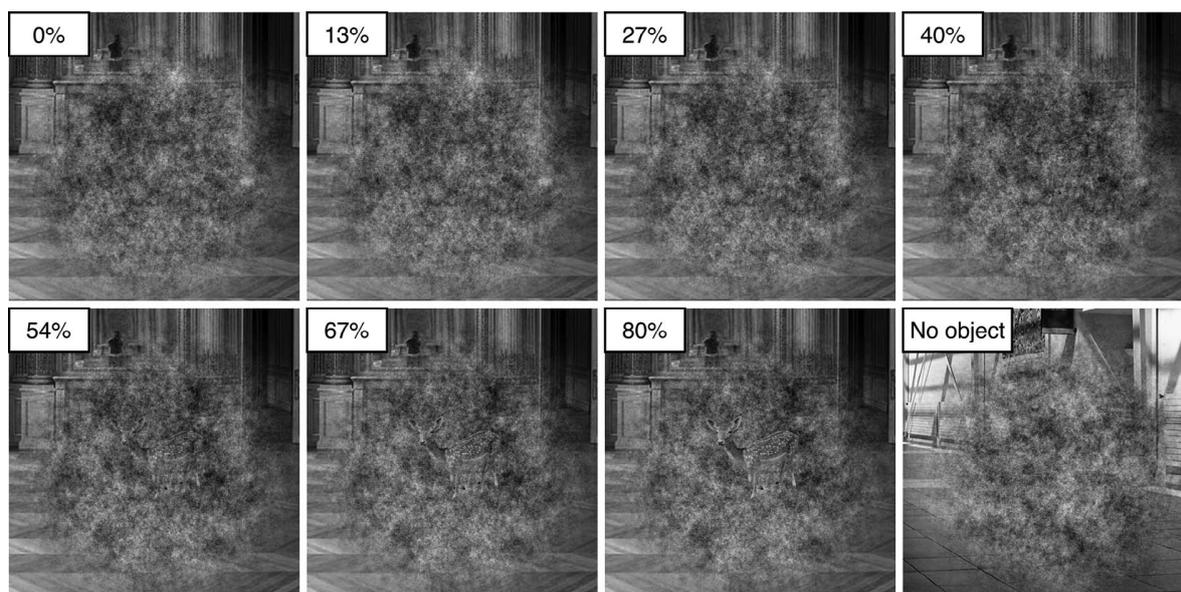


Figure 4. Portion of an indoor scene where participants fixated during the experiment, containing an object (a fawn) or no object. Participants viewed each object in seven versions, ranging from 0% to 80% of phase coherence, within 1/f noise. For illustration purposes, contrast and phase coherence of objects were slightly increased in the present figure.

alpha level of 0.05. All participants had normal or corrected-to-normal vision. They gave their informed written consent before participating in the study.

Stimuli

In the same way as for Experiment 1, we used 20 groups of four object-scene associations. Stimuli were created in MATLAB. We first gamma-corrected each object and pasted it on a 1/f noise image. We varied the visibility of the object by manipulating a phase-scrambling parameter. For each object, seven versions of the object-on-1/f image were thus created by parametrically adding coherence in the phase structure of the object, using Ales, Farzin, Rossion, and Norcia's (2012) MATLAB function. This function interpolates the phase of the Fourier transformation between the object-on-1/f image and a noise image of the same size. The seven versions ranged from 0% to 80% of phase coherence (100% being the original object on 1/f noise), increasing linearly in 13.33% steps (Figure 4). We chose to limit maximum coherence to 80% in order to minimize the ceiling performance observed in Experiment 1 with fully visible objects. Each of the seven object-on-1/f images was pasted on five different backgrounds (intact-congruent, intact-incongruent, scrambled-congruent, scrambled-incongruent, and noise-baseline). Intact scene images were gamma-corrected and equalized to obtain a mean luminance of 0.51 and mean RMS contrast of 0.24. Scrambled scene images were created by scrambling the phase of the intact scenes in the Fourier domain via random permutation. The noise

background for the baseline condition was a 1/f noise image of the same size as the scenes (i.e., 1024×768 pixels, or $24^\circ \times 18^\circ$). Object-on-1/f images were then progressively blended to the different backgrounds with a circular patch (Figure 3). Given this procedure, the circular patch had a noisy structure, allowing a more natural blending between the object and the scene compared to Experiment 1 where the circular patch was uniformly gray. We also created stimuli in which the scene backgrounds (intact and scrambled) were presented without an object (the part of the image in central vision was only filled with 1/f noise).

Procedure

The experiment was programmed on E-Prime 2.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, PA). Stimuli were displayed on the same 30-in. monitor as in Experiment 1. At a viewing distance of 50 cm, the scenes sized 24° of visual angle horizontally and 18° vertically, the larger side of the objects sized 3° , and the circular patch sized about 8° . The participant's head was supported by a chinrest in order to maintain distance. They were instructed to maintain their fixation on a little white fixation cross, centered on the screen location at which the object was presented (i.e., at the center of the screen horizontally and in the lower part vertically). The fixation cross was displayed for a random duration from 500 to 1,500 ms (on average $\sim 1,000$ ms) in order to disrupt the predictability of trials' rhythm. Each trial began with the fixation cross, followed by a first stimulus (always a background

	Congruent intact scene	Incongruent intact scene	Congruent scrambled scene	Incongruent scrambled scene	Baseline
Threshold	3.44 ± 0.65	3.76 ± 0.70	3.59 ± 0.52	3.62 ± 0.68	3.57 ± 0.51
mRT	522 ± 57	550 ± 54	541 ± 62	533 ± 58	533 ± 52
mLog(RT)	2.70 ± 0.06	2.73 ± 0.04	2.72 ± 0.05	2.71 ± 0.04	2.71 ± 0.04

Table 2. Mean threshold values, mean correct response times in milliseconds (mRT), and mean log transformed correct response times (mLog[RT]), with standard deviations, for each experimental condition (Congruence, Background).

without an object) for 150 ms, then by a second stimulus for 150 ms, and finally by a backward mask (1/f noise) for 30 ms. The second stimulus depended on the experimental condition: intact background with a congruent object, intact background with an incongruent object, scrambled background with a congruent object, scrambled background with an incongruent object, baseline 1/f noise background with an object, intact background without an object, or scrambled background without an object. The first stimulus was thus the background of the second stimulus. The interstimulus interval was 3000 ms on average. It should be noted that in the second stimulus, there was always a 1/f noise circular patch in central vision (including an object or not). Thus, in order to maintain a coherent percept in central vision throughout the trial, we also blended a 1/f noise circular patch in the background of the first stimulus.

As in Experiment 1, the task was to categorize the object as animal or furniture by pressing two keys on a keyboard. Since the visibility of the object was low in many trials, participants were encouraged to rely on their “intuition” and were instructed to respond at random when they could not see any object (without favoring one of the two response keys). Response keys were counterbalanced across participants. Accuracy and RTs were recorded at each trial. Experimental conditions were fully randomized. There were 1,720 trials: 280 intact-congruent trials (20 animal object/intact outdoor and 20 furniture object/intact indoor × 7 levels of object phase coherence), 280 intact-incongruent trials (20 animal object/intact indoor and 20 furniture object/intact outdoor × 7 levels of object phase coherence), 280 scrambled-congruent trials (20 animal object/scrambled outdoor and 20 furniture object/scrambled indoor × 7 levels of object phase coherence), 280 scrambled-incongruent trials (20 animal object/scrambled indoor and 20 furniture object/scrambled outdoor × 7 levels of object phase coherence), 280 noise-baseline trials (20 animal object/noise and 20 furniture object/noise × 7 levels of object phase coherence), 160 intact scene alone (80 intact outdoor and 80 intact indoor), and 160 scrambled scene alone (80 scrambled outdoor and 80 scrambled indoor). The experiment was split into two 1-hr experimental sessions, including pauses. Prior to completing the experimental trials, participants underwent a training

session (20 trials) to be familiarized with the stimuli and the task.

Results

Data analysis

For each participant, correct RTs were log transformed and then trimmed by removing trials for which the RT was inferior or superior to the condition average ± 2.5 SDs (0.34% of trials were excluded). Data used for the analysis of RT only included phase coherence levels for which mCR (averaged across conditions and participants) exceeded 0.75 at group level (Levels 4, 5, 6, and 7 of phase-coherence). Weibull psychometric functions were fitted to each participant’s mCR through maximum likelihood estimation for the congruent and incongruent conditions of the intact and scrambled scene conditions, as well as for the baseline condition, using the *Quickpsy* package in R (Linares & Lopez-Moliner, 2016). For each experimental condition, threshold values were derived from the psychometric functions at a mCR of 0.75 ($1 - p(\text{chance}) / 2$). Threshold values, mRT and mlog(RT), with standard deviations, for each experimental condition (congruence and background) are reported in Table 2.

Congruence effect and nature of influences: Intact versus scrambled scenes

To assess if the congruence effect was due to low-level or semantic features in the scene background, we conducted a repeated measures ANOVA on the threshold values and mLog(RT) with congruence (congruent or incongruent stimuli) and background (intact or scrambled scenes) as within-subject factors. Further pairwise comparisons were tested with two-tailed paired-sample *t* tests using a Bonferroni adjustment of alpha level to correct for multiple tests ($0.05 / 2 = 0.025$ for the two tests performed with each dependent variable). Effect size was estimated by calculating Cohen’s *d* for within subject design.

Analysis of thresholds values showed that main effects of congruence ($F[1, 17] = 4.26, p = 0.055$) and background ($F[1, 17] < 1$) were not significant but that these two factors interacted ($F[1, 17] = 7.54, p = 0.014, d = 0.65$; Figure 5). Follow-up pairwise comparisons

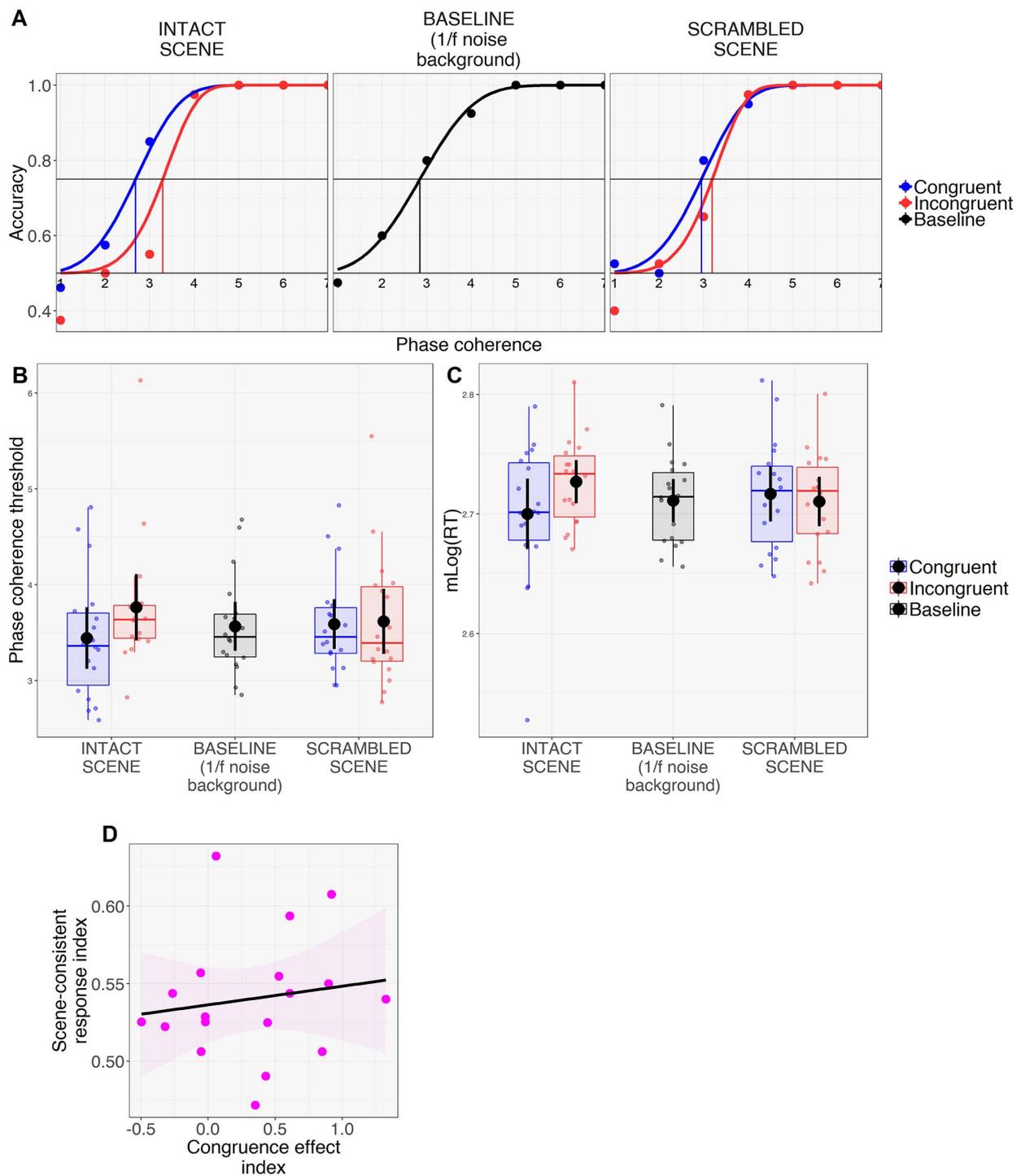


Figure 5. (A) For illustrative purposes, the figure shows a Weibull model fitted to the accuracy of one representative subject for the congruent and incongruent stimuli of the intact and scrambled background conditions, and for the baseline condition (for which the background was just 1/f noise). (B) Mean correct response rates (mCR) and (C) mean of log transformed correct RTs (mLog[RT]) plotted according to the experimental condition. Black dots and error bars indicate means and 95% CIs. Color dots are the individual observations. (D) Congruence effect index (difference in thresholds between the congruent and incongruent stimuli of the intact scene condition) as a function of the scene-consistent response index (the tendency of participants to use the category of the scene to infer the category of the object, when there is no object in the stimulus). Shading is 95% CI.

showed that thresholds were significantly lower in the congruent (3.44 ± 0.65) than in the incongruent (3.76 ± 0.70) condition for the intact scenes ($t[17] = 2.73$, $p = 0.014$, $d = 0.65$) while there was no difference between the congruent and incongruent conditions for the scrambled scenes (congruent: 3.59 ± 0.53 ; incongruent: 3.62 ± 0.68 ; $t[17] = 0.36$, $p = 0.723$).

Analysis of mLog(RT) gave very similar results with no main effect neither of congruence ($F[1, 17] = 1.56$, $p = 0.229$) nor of the background ($F[1, 17] < 1$), but there was an interaction between the two factors ($F[1, 17] = 7.13$, $p = 0.016$, $d = 0.63$; Figure 5). However, follow-up pairwise comparisons showed no main effect of congruence neither with intact scenes (congruent: 2.70 ± 0.06 ; incongruent: 2.73 ± 0.04 ; $t[17] = 2.12$, $p = 0.049$) nor with scrambled scenes (congruent: 2.72 ± 0.05 ; incongruent: 2.71 ± 0.04 ; $t[17] = 0.90$, $p = 0.383$).

These results partially support those of Experiment 1 by showing that visual information in the periphery is automatically processed and can decrease the level of perceptual quality needed to recognize objects in central vision when congruent, compared to when incongruent. However, unlike Experiment 1, we did not observe the congruence effect on RTs. Furthermore, it seems that peripheral visual information needs to carry a semantic meaning to influence the processing of the object, as no congruence effect was observed when the phase of the peripheral scene was altered (scrambled scene condition). Since there was no congruence effect with scrambled scenes, the following analyses only consider the intact scene condition.

Direction of the congruence effect: Comparison to the baseline

To evaluate the direction of the congruence effect (facilitation in the congruent condition, hindrance in the incongruent condition, or both), we compared threshold values and mLog(RT) of the baseline condition (meaningless 1/f noise background) to those of the congruent and incongruent conditions with two-tailed paired-sample t tests (alpha level Bonferroni corrected: $0.05 / 2 = 0.025$ for the two tests performed with each dependent variable). For threshold values, these tests showed no difference neither between the baseline (3.57 ± 0.51) and the congruent condition ($t[17] = 1.37$, $p = 0.190$), nor between the baseline and the incongruent condition ($t[17] = 1.99$, $p = 0.063$). For mLog(RT), there was no difference between the baseline (2.71 ± 0.04) and the congruent condition ($t[17] = 0.89$, $p = 0.388$), but participants were significantly slower in the incongruent than in the baseline condition ($t[17] = 2.72$, $p = 0.015$, $d = 0.64$).

Relation between the congruence effect and the scene-consistent response effect

In trials where no object was presented, we labeled as “scene-consistent responses” each *animal* response when an outdoor scene was presented and each *furniture* response when an indoor scene was presented. In theory, participants are supposed to respond at random between animal and furniture in the no-object trials. The proportion of scene-consistent responses thus represents the tendency of participants to use the category of the scene to infer the category of the object, when there is no object in the stimulus. We first tested if the proportion of scene-consistent responses, hereafter named *scene-consistent response index*, was different from chance level (i.e., a proportion of 0.5) in the intact condition using a one-sample t test. The proportion of scene-consistent responses averaged across participants was 0.54 ± 0.04 , and was statistically different from chance level ($t[17] = 4.28$, $p < 0.001$, $d = 1.01$), indicating that when they could not identify any object, participants’ responses were in part driven by the scene category. Then, we tested the Kendall correlation between the scene-consistent response index and the congruence effect index calculated on threshold values (incongruent *minus* congruent) of the intact scene condition. There was no correlation between these two variables ($r = 0.16$, $p = 0.36$; Figure 5). To avoid the limitation of concluding on the null hypothesis, we tested a Bayesian correlation using the *Psycho* package in R (Makowski, 2018). This Bayesian correlation analysis indicates anecdotal evidence ($BF = 1.76$) in favor of an absence of a positive association between the scene-consistent response index and the congruence effect index ($r = 0.11$, median absolute deviation = 0.21, 90% $CI [-0.23, 0.46]$). The correlation can be considered as large, moderate, small, or very small with respective probabilities of 2.94%, 15.58%, 32.78%, and 18.25%. As a reminder, the effect of the scene congruence on object categorization observed in our experiments may originate from two processes. The first would be that the implicit categorization of the scene automatically drives the response (e.g., the participant responds *animal* after having recognized an outdoor scene, independently of the recognition of the object). The second would be an *integration* of the peripheral visual input to the recognition processes occurring in central vision (i.e., the process we actually aimed to measure). In light of this, we observed no strong evidence for a link between the congruence effect index and the scene-consistent response index, suggesting that the effect of congruence on object categorization observed in our experiments was mostly due to the *integration* of the peripheral visual input to the object recognition processes in central vision, rather than being directly driven by the categorization of the scene.

Discussion

Most of the time in natural vision, eye movements direct central vision toward relevant objects, while peripheral vision continuously extracts coherent contextual information. The objective of this study was twofold: to explore how LSF extracted in peripheral vision can shape the processing of central visual inputs, and to test the predictive nature of these influences. In two independent experiments, we examined how information conveyed by LSF in peripheral vision could influence the processes leading to object categorization in central vision. Participants categorized objects surrounded by congruent information (i.e., indoor scenes for furniture, outdoor scenes for animals) or incongruent information (the other way around). Importantly in our experimental design, contextual information was restricted to peripheral vision (beyond 6° in Experiment 1 and 8° in Experiment 2) and object information to central vision. We did not filter out the HSF in peripheral vision (peripheral vision acts as a natural low-pass filter), nor did we filter out the LSF of the object in central vision (we were not interested in LSF and HSF integration at the object level). Unfortunately, we were not able to record eye movements during the experiments in order to control that participants maintained their gaze on the object location when the scene was presented. However, given that the task was to attend and categorize the object in central vision, it would have been inefficient and even demanding to make eye movements to other locations without impairing object categorization. Indeed, saccadic eye movements are usually initiated within 100–150 ms (Fischer & Weber, 1993). If participants made a saccade in the scene (presented during 150 ms in the SOA-150 condition) prior to making a new saccade toward the object (also presented during 150 ms), this puts severe constraints on (a) the time left to process visual information in the scene between a first saccade on the scene and an immediate second saccade on the object or (b) the time left to actually initiate a new saccade toward the object and categorize it. Yet, mean correct response rate was high (in Experiment 1 and in Experiment 2 when object visibility was reasonable) suggesting that eye movements did not impair object categorization in the vast majority of trials. As expected, we found an effect of the congruence of the scene in both experiments indicating that contextual information in peripheral vision can be processed automatically during a central visual field directed task. We speculate that even when irrelevant for the ongoing task, peripheral vision is used to provide information on the type of object and details that are expected to be perceived in central vision, by generating a predictive signal that controls visual processes in central vision, in line with predictive coding theories (Friston & Stephan,

2007; Lee & Mumford, 2003; Rao & Ballard, 1999) and recent models of visual recognition (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010). Moreover, in Experiment 2, the visibility threshold to accurately categorize the object was lower when the peripheral context was congruent than incongruent. In real-life conditions, where information in peripheral vision is usually congruent with central vision, predictive processes based on peripheral vision could thus boost the perceptual processing of poorly visible objects in the scene. Such mechanisms surely can be a benefit to visual recognition in low lighting and masking conditions.

Still, it is easy to overinterpret this kind of results in terms of predictive coding. We can think of another account, already discussed by Brandman and Peelen (2017), of how contextual scene information may influence object processing in this type of experimental paradigms. Without the need for predictive feedback mechanisms, object and context could be processed mostly independently, in parallel, and only later their semantic representations would be integrated. There are empirical arguments for this view. For example, the fact that visual regions in the occipito-temporal cortex are preferentially involved in the visual processing of objects (the fusiform and inferior temporal gyri; Grill-Spector, 2003; Grill-Spector, Kourtzi, & Kanwisher, 2001) and scenes (e.g., the parahippocampal cortex; Epstein, 2005; Epstein & Kanwisher, 1998) can be interpreted as the existence of partly independent representations of object and scene information. The approach adopted by Brandman and Peelen (2017) in a fMRI-MEG study to disentangle the two accounts (predictive feedback vs. parallel processing) was to pixelize the object, rendering its recognition practically impossible without a context. Behavioral results of a classification task showed that the recognition of such objects was much better when embedded in congruent scenes, than when isolated. The decoding accuracy of a classifier of BOLD activity in object and scene selective cortex was also strongly enhanced when the objects were in congruent contexts. These findings suggest that contextual information helped object categorization by shaping the *perceptual* representation of the object, via predictive feedbacks flowing through neurons of the visual cortex, and therefore are not compatible with the hypothesis of a strictly independent and parallel construction of semantic scene and object representations.

In our study, we used a different approach to test for the predictive hypothesis. In Experiment 1, we manipulated the robustness of the scene representation by presenting it for different durations before the object onset. In the object-background literature, we had no knowledge of any study having directly manipulated the sequential apparition of scene and object in order to

test for predictive aspects of object and scene integration. We speculated that the more it is possible to accumulate information in peripheral vision, the stronger the predictive signal would be and the more it would influence object recognition in central vision. We found that the effect of congruence increased linearly with scene duration, a result that is also in favor of a predictive, rather than parallel, view of background influence. In Experiment 2, mean visibility threshold for the baseline condition (noise background) was approximately in between mean threshold for the congruent condition and that for the incongruent condition (see Figure 5), even though the baseline condition did not significantly differ from either condition. For RTs, we did observe a difference between the incongruent condition and the baseline. Overall, the pattern of results is not clear-cut and, as it stands, rather suggests that an incongruent context delayed the object categorization. Therefore, we did not replicate the facilitation effect observed with congruent background by Brandman and Peelen (2017) with poorly visible object. Nonetheless, the control condition in their study was simply the object passed onto a uniform background. We rather embedded the object in a noise background following the relationship between amplitude and spatial frequency that is typical of natural scenes ($1/f^2$; Ruderman & Bialek, 1994). It is thus possible that the mere presence of noise in peripheral vision boosted the perception of the object, for example via mechanisms of stochastic resonance (McDonnell & Ward, 2011).

A substantial problem with sequential presentation is that we cannot be sure about the mechanisms responsible for a given participant's response. As mentioned above, a given object categorization can be (a) "overhasty," i.e., driven directly (and erroneously in incongruent trials) by the recognition of the scene category, or (b) actually driven by the integration of the scene to the object representation. In Experiment 2, we wanted to ensure that the congruence effect was due to the second account. We included trials in which no object was present, while participants were not aware of the manipulation. Looking at behavioral responses to those stimuli, we measured by how much participants had a tendency to use the category of the scene to infer the category of the object even though there was no object in the stimulus. First, we found that participants' responses were indeed driven by the scene category ($0.54\% \pm 0.04\%$ of responses were consistent with the scene category, significantly different from chance). In a predictive coding framework, related effects have been shown, where feedback activity carrying information about the surrounding context in the visual cortex was measured without any feedforward input (Smith & Muckli, 2010). However, we found that the tendency to respond based on the scene

category did not correlate with the size of the congruence effect, suggesting that the latter account (b) was more likely to drive our results compared to the former account (a).

The other aim of our study was to characterize further the nature of the peripheral influence. We manipulated the presence of a semantic content in the background scene by presenting it either in its intact version or in a scrambled version made by disrupting the phase of the image. We found no effect of congruence when the phase of the scene was scrambled, suggesting that the mere low-level attributes of scenes do not influence object categorization.¹ In fact, this would make sense if we think of situations where some artificial objects are typically encountered in natural environments (e.g., benches in public gardens) or natural objects in artificial environments (e.g., a cat in a living room). In such situations, the low-level visual information (i.e., the amplitude spectrum) available in peripheral vision would not be consistent with that of the object. Thus, our results suggest that predictions rather contain more mid- or high-level information and that they are initiated in a higher level cortex (for example, in the orbitofrontal cortex or in regions of the ventral pathway) that represents semantics aspects of the environment, as postulated by some models of visual recognition (Bar, 2003; Fabre-Thorpe, 2011; Kauffmann, Bourgin, et al., 2015). Consistently, studies by Loschky and collaborators (Loschky & Larson, 2008; Loschky et al., 2007) have shown that the Fourier amplitude spectrum is insufficient for scene gist recognition. This suggests that the scene must be recognized at some level (e.g., coarse gist recognition) in order to influence the object processing. The mere low-level signal would not be sufficient. Yet, it is difficult to conclude on the null hypothesis. It is thus possible that the physical information contained in the scrambled scenes alone does influence object processing at a lower extent, but that we were not able to measure this effect here. For example, this can be due simply to a lack of statistical power, since we planned our effect size for Experiment 2 based on the congruent effect of Experiment 1, where only intact scenes were presented. It is also possible that phase scrambling is not a good model of low-level predictive processes in the brain, and that the spatial distribution of low-level features—available in the phase spectrum—is in fact important for such processes. Stojanoski and Cusack (2014) have compared the neural activity at the earliest stages of the visual system induced by intact images, phase-scrambled images, as well as diffeomorphic images (meaningless images resembling images printed on a distorted rubber sheet). They have shown that these regions do not respond to phase-scrambled stimuli in the same way as they do to intact ones. On the contrary, neural activity induced by diffeomorphic

images is indistinguishable to that induced by intact images. Diffeomorphic stimuli may therefore be a better way to model low-level visual processes.

All that being said, it seems important to highlight some matters regarding the effect of congruence in our experiments. Initially, we hypothesized that low-resolution information in peripheral vision could be used to help perceptual processing of the object in central vision via predictive processes. However, the effect of congruence was not observed when the scene was simultaneously presented with the object (SOA-0 condition), nor when the scene was presented 30 ms before the onset of the object (SOA-30 condition). Yet, according to spatial frequency-based models of visual recognition, LSF are processed faster than HSF and thus could influence the processing of the latter when perceived simultaneously. It seems that the visual system must have enough time to process peripheral LSF for it to influence object recognition (here, in the SOA-150 condition). Second, in the SOA-150 condition, the difference in RT between congruent and incongruent conditions was only of 20 ms, as pointed out before. Third, in Experiment 2, the congruence effect on RT observed in Experiment 1 did not replicate (participants were not significantly faster to categorize the object in the congruent than incongruent condition), although we observed the effect on thresholds.

Therefore, it is possible that peripheral vision only plays a moderate role in the predictive processes leading to object recognition, and that LSF available in central vision are preferentially used to initiate such predictive processes. Nonetheless, it should be noted that our task was a superordinate (animal vs. furniture) categorization task. A basic level categorization task could benefit more from low-resolution contextual information in peripheral vision. Future studies should compare directly LSF based predictions from central vision to those from peripheral vision, and also investigate other levels of object categorization (e.g., basic level). On the other hand, it is possible that LSF are not a good descriptor of peripheral vision. For example, crowding and ensemble perception are important properties of peripheral vision that we have not considered here. Influential models of peripheral vision such as Balas, Nakano, and Rosenholtz (2009) suggest that such mechanisms lead to consider peripheral vision as qualitatively different from central vision. Peripheral vision would represent summary statistics of different features of the environment (orientations, size, textures, hue) within pooling regions. Therefore, summary statistics could be a better descriptor of peripheral vision than LSF.

To conclude, this study shows that low-resolution information in peripheral vision can be unintentionally processed and integrated to information in central vision, if available for a sufficient amount of time. This

effect is not likely to be due to response strategy, but rather on mechanisms of periphery-based predictions integrated during the processing of local elements in central vision. The study also suggests that predictions rely on the semantic processing of peripheral information, since mere low-level information did not influence the processing of the object in central vision. Such mechanisms could boost perception of poorly visible objects. Given that the effects are rather small and unstable, further studies manipulating the level of object categorization and comparing predictions coming from central and peripheral vision are needed to explore the issue further.

Keywords: peripheral vision, scene, context effect, spatial frequency, predictive coding

Acknowledgments

Alexia Roux-Sibilon was supported by the “Alpes Grenoble Innovation Recherche” grant from the pole “Chimie-Biologie-Santé” of University Grenoble Alps (AGIR-POLE CBS). Louise Kauffmann was supported by the NeuroCoG IDEX UGA (Initiatives D’EXcellence project of University Grenoble Alps) in the framework of the *Investissements d’avenir* program from the Agence Nationale de la Recherche (ANR-15-IDEX-02). We thank Valerie Goffaux and Kevin Parisot for their help with the stimuli.

Commercial relationships: none.

Corresponding author: Alexia Roux-Sibilon.

Email: alexia.roux-sibilon1@univ-grenoble-alpes.fr.

Address: University Grenoble Alps, Univ. Savoie Mont Blanc, CNRS, LPNC, Grenoble, France.

Footnote

¹ It should be noted that the absence of a congruence effect in the scrambled condition does not preclude that the mere presence of noise in this condition may have boosted the perception of the object, as we previously suggested for the baseline condition (1/f noise background).

References

Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of Vision*, 12(10):

- 18, 1–18, <https://doi.org/10.1167/12.10.18>. [PubMed] [Article]
- Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience*, *29*(34), 10638–10652.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18, <https://doi.org/10.1167/9.12.13>. [PubMed] [Article]
- Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision*, *16*(2):9, 1–14, <https://doi.org/10.1167/16.2.9>. [PubMed] [Article]
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.
- 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., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.
- Bar, M., & Ullman, S. (1993). Spatial context in recognition. *Perception*, *25*(3), 343–352.
- Boucart, M., Moroni, C., Szaffarczyk, S., & Tran, T. H. C. (2013). Implicit processing of scene context in macular degeneration. *Investigative Ophthalmology and Visual Science*.
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, *37*(32), 7700–7710.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, *300*(1), 5–25.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, *292*(4), 497–523.
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559.
- Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*(6), 954–978.
- Epstein, R., & Kanwisher, N. (1998, April 9). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 243.
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(3), 553–567.
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458.
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422.
- Hegd e, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439.
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13):11, 1–18, <https://doi.org/10.1167/8.13.11>. [PubMed] [Article]
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297.
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394–2405.
- Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes: A dynamic causal modeling study. *Brain and Cognition*, *99*, 46–56.
- Kauffmann, L., Ramano el, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37.

- Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* (pp. 221–244). New York, NY: Springer.
- 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.
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America, A*, 20(7), 1434–1448.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience*, 4(5), 533–539.
- Linares, D., & Lopez-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, 8(1), 122–131.
- Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, 8(1):4, 1–9, <https://doi.org/10.1167/8.1.4>. [PubMed] [Article]
- Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1431–1450.
- Makowski, D. (2018). The Psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 3(22), 470.
- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176–184.
- McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, 12(7), 415.
- Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, 186, 103–112.
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., . . . Vuilleumier, P. (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.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Rémy, F., Saint-Aubert, L., Bacon-Macé, N., Vayssière, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: A life-span study. *Vision Research*, 91, 36–44.
- Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. In *Advances in Neural Information Processing Systems* (pp. 551–558).
- 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.
- Shams, L., & Von Der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, 42(22), 2547–2554.
- Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, 8(1), 547–581.
- Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences*, 107(46), 20099–20103.
- Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, 14(12):6, 1–16, <https://doi.org/10.1167/14.12.6>. [PubMed] [Article]
- Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., & Teder, W. A. (2011). Contextual influences on rapid object categorization in natural scenes. *Brain Research*, 1398, 40–54.
- Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., . . . Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, 115(35), 8835–8840.
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, 8(1), 10853.
- Van Essen, D., & DeYoe, E. A. (1995). Concurrent processing in the primate visual cortex. In *The Cognitive Neurosciences* (pp. 383–400).
- Woodhead, Z. V. J., Wise, R. J. S., Sereno, M., & Leech, R. (2011). Dissociation of sensitivity to spatial frequency in word and face preferential areas of the fusiform gyrus. *Cerebral Cortex*, 21(10), 2307–2312.
- Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, 24(11), 2248–2261.

Annexe 3

RESEARCH ARTICLE

Scene and human face recognition in the central vision of patients with glaucoma

Alexia Roux-Sibilon¹*, Floriane Rutgé², Florent Aptel², Arnaud Attye³, Nathalie Guyader⁴, Muriel Boucart⁵, Christophe Chiquet², Carole Peyrin¹

1 Univ. Grenoble Alpes, CNRS, LPNC, Grenoble, France, **2** Department of Ophthalmology, University Hospital, Grenoble, France, **3** Department of Neuroradiology and MRI, University Hospital, Grenoble, France, **4** Université Grenoble Alpes, CNRS, GIPSA-Lab UMR 5210, Grenoble, France, **5** Université de Lille, CNRS, SCALab UMR 9193, Lille, France

* These authors contributed equally to this work.

* alexia.roux-sibilon1@univ-grenoble-alpes.fr



Abstract

Primary open-angle glaucoma (POAG) firstly mainly affects peripheral vision. Current behavioral studies support the idea that visual defects of patients with POAG extend into parts of the central visual field classified as normal by static automated perimetry analysis. This is particularly true for visual tasks involving processes of a higher level than mere detection. The purpose of this study was to assess visual abilities of POAG patients in central vision. Patients were assigned to two groups following a visual field examination (Humphrey 24–2 SITA-Standard test). Patients with both peripheral and central defects and patients with peripheral but no central defect, as well as age-matched controls, participated in the experiment. All participants had to perform two visual tasks where low-contrast stimuli were presented in the central 6° of the visual field. A categorization task of scene images and human face images assessed high-level visual recognition abilities. In contrast, a detection task using the same stimuli assessed low-level visual function. The difference in performance between detection and categorization revealed the cost of high-level visual processing. Compared to controls, patients with a central visual defect showed a deficit in both detection and categorization of all low-contrast images. This is consistent with the abnormal retinal sensitivity as assessed by perimetry. However, the deficit was greater for categorization than detection. Patients without a central defect showed similar performances to the controls concerning the detection and categorization of faces. However, while the detection of scene images was well-maintained, these patients showed a deficit in their categorization. This suggests that the simple loss of peripheral vision could be detrimental to scene recognition, even when the information is displayed in central vision. This study revealed subtle defects in the central visual field of POAG patients that cannot be predicted by static automated perimetry assessment using Humphrey 24–2 SITA-Standard test.

OPEN ACCESS

Citation: Roux-Sibilon A, Rutgé F, Aptel F, Attye A, Guyader N, Boucart M, et al. (2018) Scene and human face recognition in the central vision of patients with glaucoma. PLoS ONE 13(2): e0193465. <https://doi.org/10.1371/journal.pone.0193465>

Editor: Mark Alexander Williams, Macquarie University, AUSTRALIA

Received: August 8, 2017

Accepted: February 12, 2018

Published: February 26, 2018

Copyright: © 2018 Roux-Sibilon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data files are available from the public repository <https://osf.io/79r2b/#>.

Funding: ARS was supported by University Grenoble Alpes PhD grant "AGIR-POLE-CSB", <https://www.communaute-univ-grenoble-alpes.fr/presentation/annonces/agir-pole-634391.htm#CBS>. FR was supported by Fondation de France grant "Berthe Fouassier", <https://www.fondationdefrance.org/fr/fondation/fondation-berthe-fouassier>. The funders had no role in study

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Primary open-angle glaucoma (POAG) is an ocular disease characterized by a progressive destruction of retinal ganglion cells and the optic nerve. This results in progressive vision loss affecting mainly peripheral vision [1]. However, this visual field loss is rather different than the black tunnel vision that patients are believed to experience. Vision loss includes missing parts and blurred distortions [2,3]. As POAG develops slowly and without pain, the diagnosis is often made at an advanced stage. By this time the visual handicap is irreversible. Therefore, it is important to develop and propose visual tests and tools to diagnose the disease at early stages. Then disabilities can be quantified at different stages, to identify individuals who might benefit from rehabilitative intervention. As the disease is due to the destruction of retinal ganglion cells, studies of POAG usually focus on low-level aspects of visual processing, i.e. retinal sensitivity to visual stimulation. The visual field defect is mostly assessed by static automated perimetry. This is based on luminance increment detection of small dots and allows evaluation of retinal sensitivity. However, POAG patients often report discomfort in their daily life, even at stages when static automated perimetry only indicates a slight peripheral vision loss [4]. Furthermore, the current state of our understanding of visual cognition implies that simple aspects of visual information are integrated into more complex percepts. Retinal sensitivity loss may thus impact more complex aspects of visual processing in POAG patients, even early in the disease. Converging descriptions of complex visual deficits in glaucoma include form and motion extraction [5], face and object recognition [6,7], reaching and grasping [8], mobility [4], executions of natural actions [9]. Also, specific eye movement patterns have been observed during reading [10], face recognition [11], exploration of scenes [12,13,14], visual search of objects [15].

Importantly, there is also evidence of complex visual impairment, even in areas of the visual field classified as normal by static automated perimetry [5,7]. Usually POAG causes a decrease in contrast sensitivity, correlated with the visual field defect [16]. However, Lenoble et al. [7] showed that people diagnosed with glaucoma, without central visual field loss as reflected by static automated perimetry, performed worse than controls in categorizing single objects presented foveally at diminished luminance contrast of stimuli. These results suggest that recognition abilities of patients may be more impacted than mere retinal sensitivity (i.e. detection of small dots).

The aim of the present study was to assess the visual abilities of POAG patients with respect to the perception of scene photographs (Scene experiment) and human face photographs (Face experiment) in the central visual field, a region considered to be relatively preserved and only affected in the later stages of the disease. The critical manipulation in these experiments was the level of visual processes required to perform a task, while presenting the same visual stimuli. For this purpose, we first evaluated patients' retinal sensitivity using a detection task of photographs of scenes or human faces, with instructions mimicking those of static automated perimetry. We then evaluated the recognition of these photographs with a categorization task. Primordially, we tested the participants' ability to detect and categorize the same stimuli, which enabled us to directly compare performances in detection (low-level visual processes) and categorization (higher-level visual processes). We considered the difference in performance between detection and categorization to be a reliable indicator of the cost of higher-level visual processes. POAG patients were divided into two groups: patients with a partially affected central visual field, and those with a preserved central visual field, as observed by perimetry using Humphrey 24–2 SITA-Standard test. Performances of the two groups of patients were compared to age-matched normally sighted participants. According to static automated perimetry results, we expected that only patients with a central defect (CD) would have altered

capacities for low-level processes (the detection task), with respect to controls. Further, we proposed that the peripheral field loss would impact high-level functions in central vision in a way not predicted by the Humphrey 24–2 SITA-Standard test. We expected a categorization deficit, with respect to controls, not only for patients with a central visual field defect but also for those with no central defect (NoCD). Additionally, we considered the impact of luminance contrast of the stimuli. Two low levels were presented (10% and 2.5%), reproducing the situation of twilight or a darkened room. According to the results of Lenoble et al. [7], we expected that the difference of performance between the Detection and the Categorization tasks would be larger for the lowest contrast level (2.5%), and that this interaction would be stronger for NoCD patients than for controls.

Methods

Participants

Twenty-two subjects (12 women) currently being treated with topical therapy for POAG were included in the experiment. Diagnosis was established in the ophthalmologic department of the Grenoble Alpes University Hospital. Patients participated in this non-interventional behavioral experiment following their clinical examination. The best corrected visual acuity in the eye to be tested was at least 6.3/10 on a Monoyer scale for distance vision (or 0.2 logMAR) and about 1.5 or 2 Parinaud for near vision. If both eyes corresponded to these criteria, the tested eye was chosen to balance the two POAG groups (with or without a central visual field defect). Monocular assessment of POAG patients' visual field was performed with adapted corrected lenses, through a Humphrey Visual Field Analyzer (Carl Zeiss, Meditec, Dublin, CA; 24–2 SITA-Standard procedure) in each eye. The quality of the patients' central visual field was examined based on the individual deviation probability plot of the four central points tested, which cover the central 6° of the visual field (3° in each hemifield). A deficit in the central visual field was defined by at least one of the four points presenting a probability inferior of 2% of being normal. This classification resulted in two POAG groups (Fig 1); 11 patients with a CD (5 women), and 11 NoCD patients (7 women). Mean visual acuity for CD patients was 9.6/10 (± 0.9) Monoyer, and 2 Parinaud for all patients. Mean visual acuity for NoCD patients was 9.5/10 (± 1.0) Monoyer, and 2 Parinaud for all patients. Twenty-five control subjects (14 women) were also tested unilaterally, in the eye with the best corrected visual acuity (Control group). Mean visual acuity for the Control group was 10/10 Monoyer and 2 Parinaud for all participants. Control subjects also underwent monocular automatized Humphrey visual field assessment with a 24–2 SITA-Standard procedure in each eye to ensure the absence of campimetric deficit.

The three groups were approximately age-matched (Control group: 68.0 ± 6 years, range 54–83 years; CD group: 67.9 ± 10 years, range 52–79 years; NoCD group: 67.1 ± 9 years, range 55–79 years; *t*-tests: Control vs. CD: $t(34) = 0.03$, $p = 0.97$; Control vs. NoCD: $t(34) = 0.31$, $p = 0.76$; CD vs. NoCD: $t(20) = 0.18$, $p = 0.86$). A complete ophthalmological examination (best corrected visual acuity, Goldmann tonometry, anterior segment, and fundus examination) was performed for all participants. None exhibited cognitive impairments (mini mental state examination score $> 24/30$). Participants with psychiatric, neurological, and ocular disorders (age-related macular degeneration, cataract except for uncomplicated cataract surgery) were not included in the study, and all participants gave their informed written consent before participation. The study was carried out in agreement with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the local ethics committee (CERNI, COMUE Université Grenoble Alpes, IRB00010290).

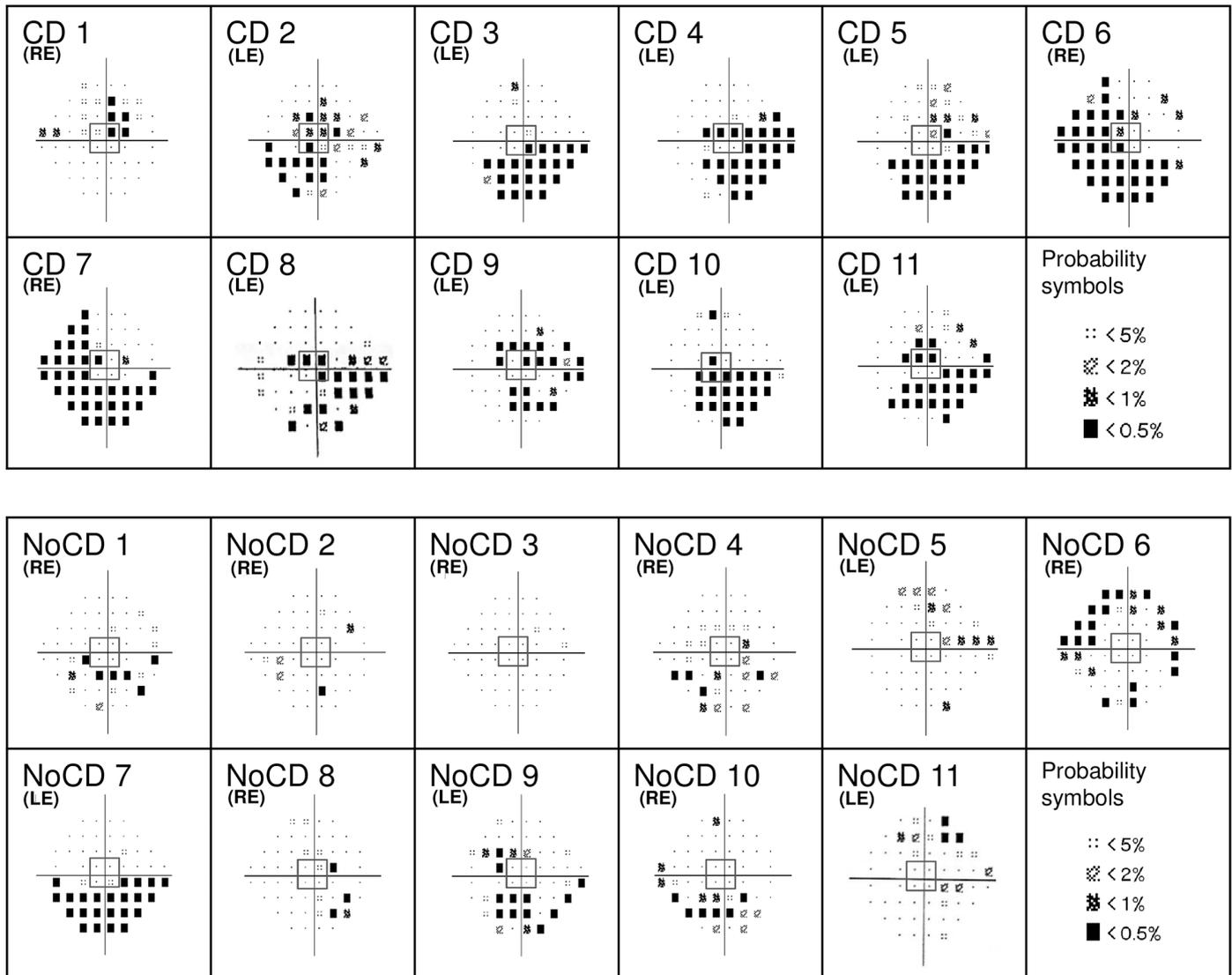


Fig 1. Individual mean deviation probability plots for patients with primary open-angle glaucoma. A deficit in the central visual field was defined by at least one of the four central points (framed in central grey squares) presenting a probability of less than 2% of being normal. The tested eye is indicated for each patient (LE: left eye; RE: right eye). CD: Patients with central defect; NoCD: Patients with no central defect.

<https://doi.org/10.1371/journal.pone.0193465.g001>

Stimuli and procedure

All participants performed two successive experiments, a Scene experiment and a Face experiment (Fig 2). Stimuli were elaborated using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA). Stimuli in the Scene Experiment were selected from previous research [17,18]. These were gray-scale photographs of 20 scenes (192 × 144 pixels) belonging to two semantic categories: indoors (kitchens, living-rooms) and outdoors (streets, buildings, houses). Both categories were equivalent in terms of visual cluttering and all scenes had a similar energy distribution in both spatial frequency and dominant orientations (similar mean amplitude spectrum). Thus, categorization based on physical properties of the images were avoided. Stimuli in the Face experiment, selected from previous research [19],

(a) Scene experiment



(b) Face experiment

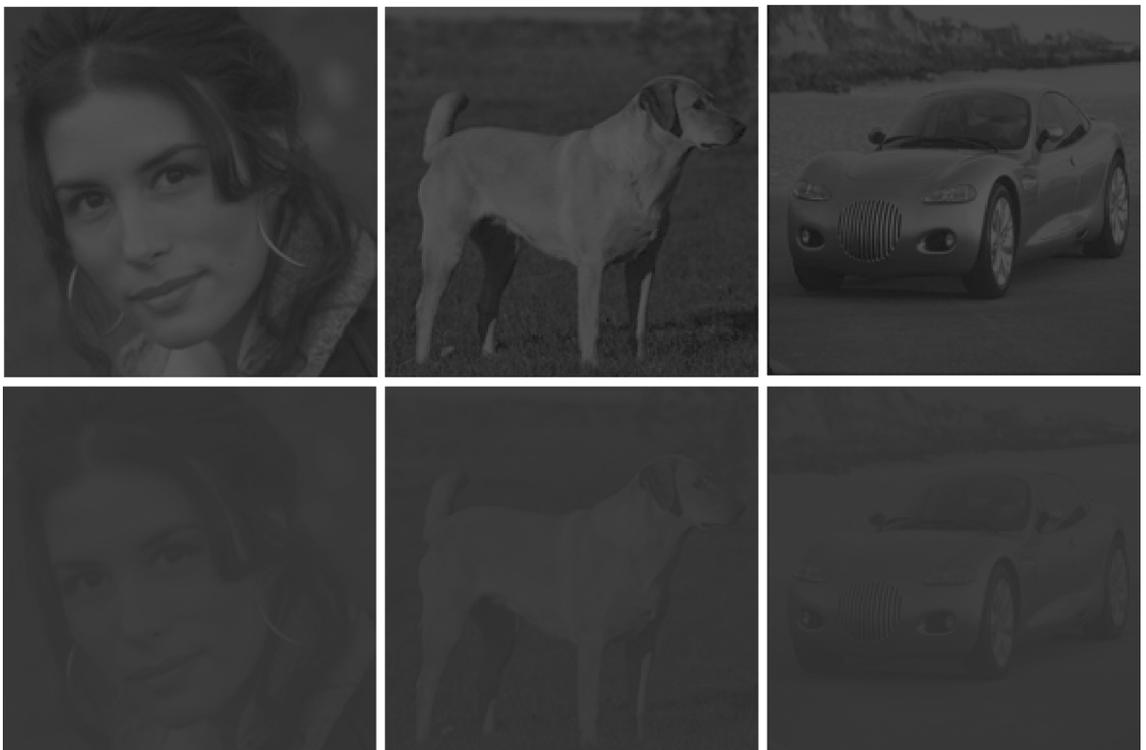


Fig 2. Examples of stimuli for (a) Scene and (b) Face experiments. Images were presented at two contrast levels (10% and 2.5%). For illustrative purposes, the contrast level of the images was slightly increased.

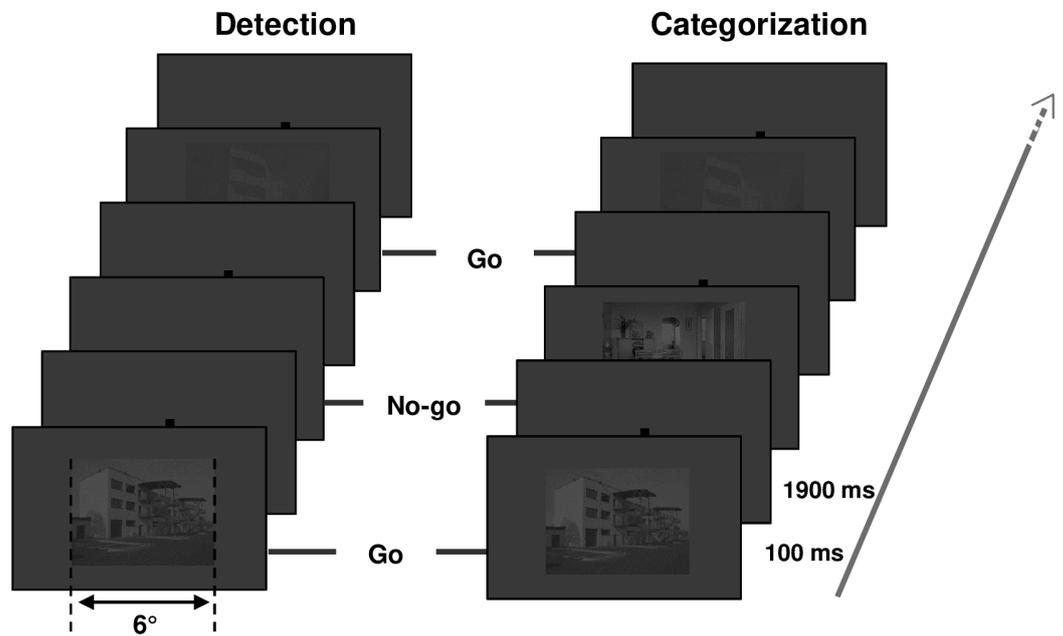
<https://doi.org/10.1371/journal.pone.0193465.g002>

were gray-scale photographs of 10 human faces, five animals, and five vehicles (192×192 pixels). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) on a 30" monitor, with a resolution of 2560×1600 pixels, at a refreshing rate of 60 Hz and with a viewing distance of 38 cm. Optical correction was used for participants requiring visual correction at this distance. Stimuli covered the central 6° of the visual angle width at this viewing distance. Hence, the region covered by the four central points of the automatized visual field evaluation used to determine whether POAG patients presented a central defect was matched. To maintain the distance and central position, participants' heads were supported by a chinrest. We measured the gamma function of our monitor (i.e. the luminance of the display for different values of uniform gray level stimuli generated with Matlab) using a calibration tool (Spyder5Elite). We then applied a gamma-correction to each stimulus to linearize the gamma function. Decreasing luminance contrast of the images had a proportional effect on the low (dark shades) and high (bright shades) luminance values. We then created two contrast level versions of each image, by manipulating the Root-mean-square contrast (RMS; standard deviation of the luminance [20]). This manipulation resulted in one version of the image with a luminance contrast level of 2.5% and one version with a luminance contrast level of 10% (RMS contrast of respectively 0.007 and 0.03 for luminance values between 0 and 1; see Fig 2). The mean luminance values of all images as well as the background luminance of the screen were set at 72 on a 256 gray-level scale (0.28 for luminance values between 0 and 1; 10 cd/m^2) to match the luminance of the Humphrey Visual Field Analyzer background (luminance of 31.5 apostilbs, corresponding to approximately 72 gray-levels or 10 cd/m^2).

Each experiment (Scene and Face) consisted of two experimental blocks dedicated to two different tasks: a Detection task and a Categorization task (Fig 3). For both experiments, the Detection task was to press the keyboard spacebar when a stimulus (a scene or a human face) appeared on the screen, based on a go/no-go trial paradigm. The Detection task consisted of 20 go trials and 20 no-go trials. In go trials, an image (a scene or a face) was presented against the uniform grey background (10 trials for each contrast level). In no-go trials, there was no image, only the uniform grey background. Participants were explicitly told not to try to recognize the image. However, they had to respond as quickly as possible, as soon as they could see the presence of an image. For the Categorization tasks, participants had to press the keyboard spacebar when the image belonged to a target category, but not when it belonged to a distractor category. Target and distractor categories for the Scene experiment were counterbalanced across participants (either indoors as target and outdoors as distractors, or the inverse). In the Face experiment, the target category was always faces and distractors were always animals or vehicles. Participants were told to respond as quickly as possible, as soon as they could see an image belonging to the target category. The Categorization task consisted of 20 go trials (target category, 10 for each contrast level) and 20 no-go trials (distractor categories, 10 for each contrast level). In each experiment, go trial stimuli were the same in the detection and categorization tasks; only task demands changed.

A stimulus (either a photograph or nothing for no-go trials of the Detection block) was displayed against the uniform grey background for 100 ms to allow a sufficient processing time while minimizing saccades. A central fixation point was displayed against the uniform grey background before and after the trial stimulus (an image against the grey background or only the grey background) to control the gaze direction to the center of the screen. The inter-

(a) Scene experiment



(b) Face experiment

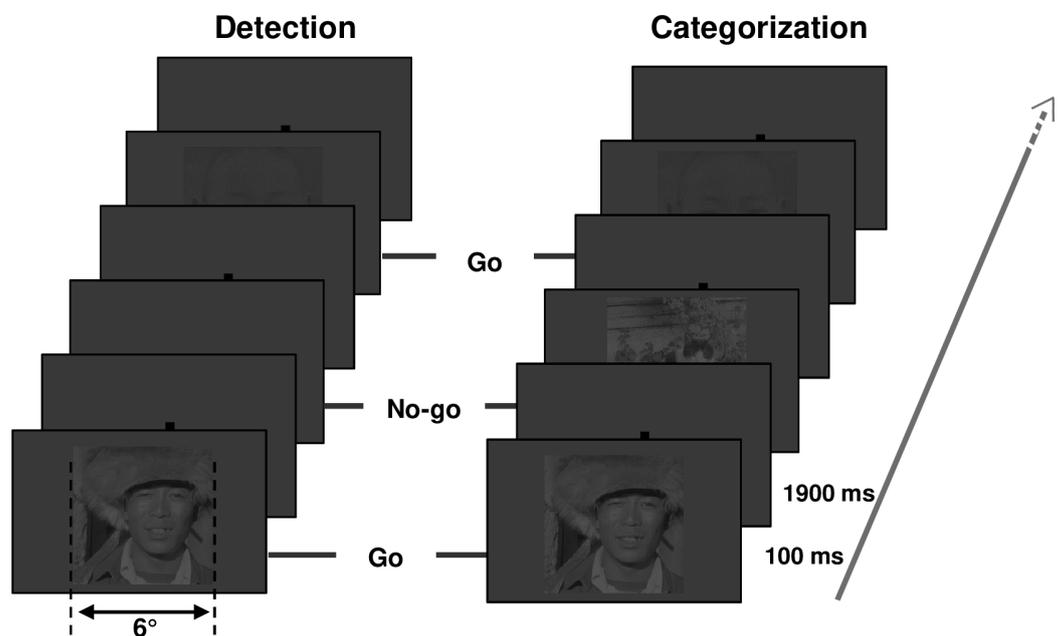


Fig 3. Schematic procedure for (a) Scene and (b) Face experiments. For each experiment, examples of two go trials and two no-go trials are represented by task (Detection and Categorization). Visual stimulation for go trials was the same for the two tasks. Images were presented in two contrast levels (10% and 2.5%). For illustrative purposes, the contrast level of the images was slightly increased.

<https://doi.org/10.1371/journal.pone.0193465.g003>

stimulus interval was 1900 ms. Scene and Face experiments were counterbalanced across participants of each group. In each experimental session, the Detection task always preceded the Categorization task, to avoid contamination of the Detection task by categorization processes. Contrast conditions in each block were presented in random order. Before each experimental block, participants underwent a short training block comprising 12 trials using stimuli that differed from the ones used in the experiment. The participant's response was recorded for each trial (either a *go*-response or a *no-go*-response) and reaction times were recorded for the *go* responses.

Data analysis

The participants' responses were designed for each trial as follows: correct detection of a target (an image in the Detection task or a specific category in the Categorization task) was a hit, while the absence of a response when a target was present was an omission. Furthermore, a detection of a target when no target was presented was a false alarm and no response when no target was presented was a correct rejection. Based on the proportion of hits, omissions, false alarms, and correct rejections, a d' index of sensitivity was calculated for each participant and each experimental condition, as part of the Signal detection theory. One analysis of variance for the Scene experiment and one for the Face experiment were then carried out on the d' index of sensitivity using Statistica 10.0 software (Statsoft, Tulsa, USA). The between-subject factors were Group (Control, CD, and NoCD), and within-subject factors were Task (Detection and Categorization) and Luminance contrast level (2.5% and 10%). Effect sizes were estimated by partial eta-squared (η^2). The significance level of tests was set at 0.05. Due to the presence of extreme proportions of omission for several patients, especially for the Categorization task when the luminance contrast was 2.5%, the mean correct reaction times could not be calculated. Thus, no analysis was conducted on reaction times.

Results

Scene experiment

The analysis of variance performed on d' first revealed a main effect of group ($F(2,44) = 15.40$, $p < 0.001$, $\eta^2 = 0.41$) with a higher sensitivity for controls than for CD patients (controls: 3.24 ± 0.44 ; CD: 2.68 ± 0.68 ; $F(1,44) = 33.52$, $p < 0.001$) and for NoCD patients (controls: 3.24 ± 0.44 ; NoCD: 2.10 ± 0.74 ; $F(1,44) = 8.82$, $p < 0.05$). The main effect of the task was significant ($F(2,44) = 176.34$, $p < 0.001$, $\eta^2 = 0.80$), with a lower sensitivity for the Categorization task (2.20 ± 1.02) than for the Detection task (3.48 ± 0.60). Furthermore, the Task interacted with the Group factor (Fig 4A; $F(2,44) = 8.27$, $p < 0.001$, $\eta^2 = 0.27$). Planned comparisons revealed that, for the Detection task, CD patients had a lower sensitivity than controls (CD: 3.06 ± 0.92 ; controls: 3.71 ± 0.33 ; $F(1,44) = 10.74$, $p < 0.001$), while the difference between NoCD patients and controls was not significant (NoCD: 3.40 ± 0.48 ; controls: 3.71 ± 0.33 ; $F(1,44) = 2.38$, $p = 0.13$). For the Categorization task, mean comparisons showed that CD patients, and notably NoCD patients, had a lower sensitivity than controls (CD: 1.14 ± 0.78 ; NoCD: 1.96 ± 0.98 ; controls: 2.78 ± 0.67 ; CD vs. controls: $F(1,44) = 33.71$, $p < 0.001$; NoCD vs. controls: $F(1,44) = 8.45$, $p < 0.01$). Considering only controls and CD patients as groups, the Task interacted with the Group suggesting that the difference between sensitivity for the Categorization task and for the Detection task was larger for CD patients than for controls (CD patients: 1.44 ± 0.73 ; controls: 0.93 ± 0.60 ; $F(1,44) = 15.73$, $p < 0.001$). Importantly, considering only controls and NoCD patients as groups, the Task also interacted with the Group suggesting that the difference between sensitivity for detection and categorization

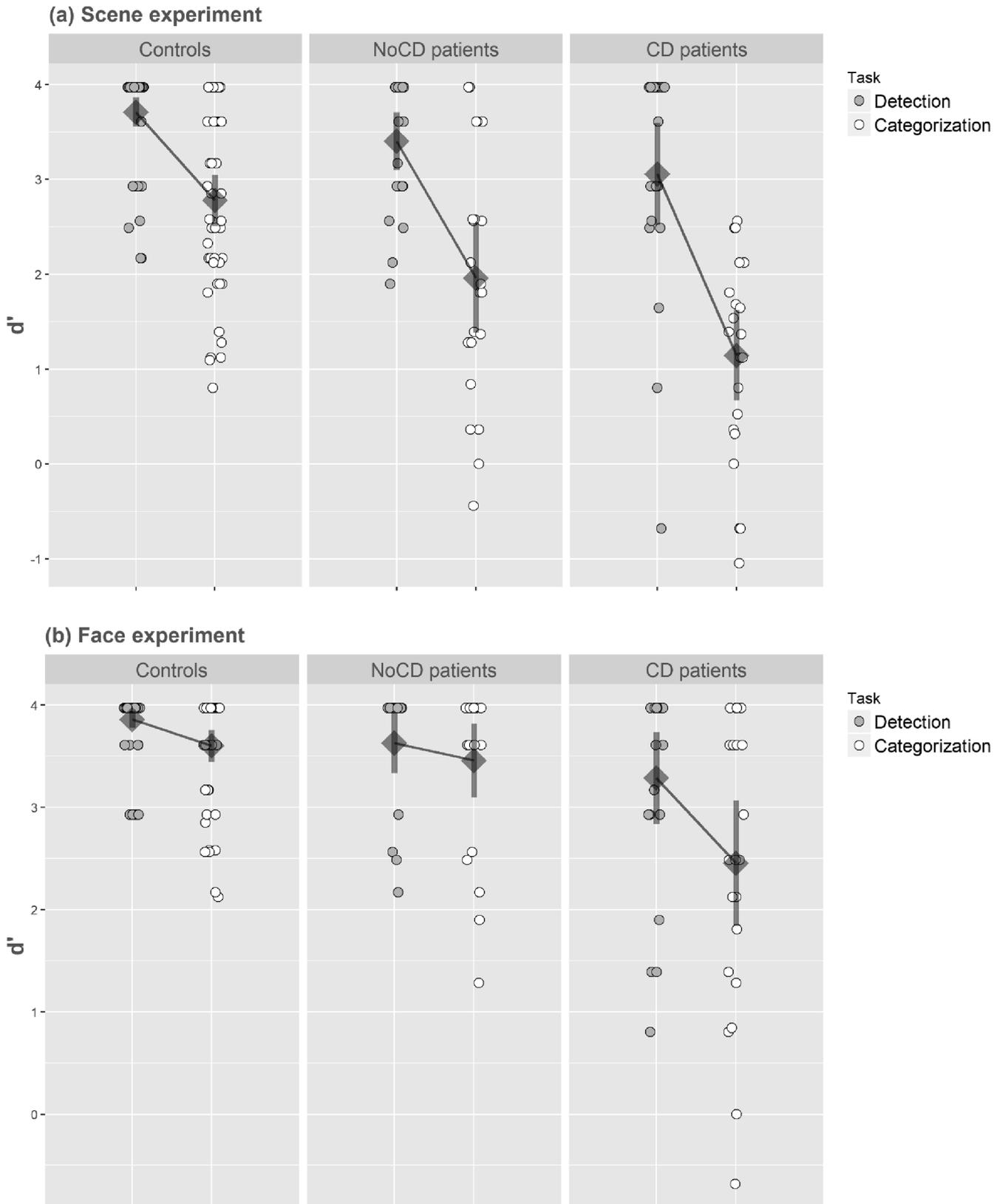


Fig 4. Index of sensitivity (d') in (a) the Scene experiment and (b) the Face experiment. Results are presented as a function of the group (primary open-angle glaucoma patients with central defect: CD; primary open-angle glaucoma patients without central defect: NoCD; and normally sighted age-matched participants: controls), and the task (detection vs. categorization). Error bars correspond to 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0193465.g004>

was also larger for NoCD patients than for controls (NoCD patients: 1.91 ± 0.82 ; controls: 0.93 ± 0.60 ; $F(1,44) = 4.26$, $p < 0.05$).

There was no main effect of the Luminance contrast level ($F(2,44) = 1.89$, $p = 0.16$) but the interaction between Contrast level \times Task \times Group was significant (Fig 5; $F(2,44) = 3.96$, $p < 0.05$, $\eta^2 = 0.15$). When the luminance contrast of scenes was 10%, planned comparisons revealed that, for the Detection task, there was no difference in sensitivity between controls and NoCD patients (controls: 3.91 ± 0.28 ; NoCD: 3.68 ± 0.45 ; $F(1,44) = 2.94$, $p = 0.09$), nor between controls and CD patients (controls: 3.91 ± 0.28 ; CD: 3.71 ± 0.52 ; $F(1,44) = 2.19$, $p = 0.15$). For the Categorization task, mean comparisons showed that CD patients as well as NoCD patients had a lower sensitivity than controls (CD: 1.70 ± 0.82 ; NoCD: 2.71 ± 0.93 ; controls: 3.35 ± 0.69 ; CD vs. controls: $F(1,44) = 33.94$, $p < 0.001$; NoCD vs. controls: $F(1,44) = 5.00$, $p < 0.05$). When the luminance contrast of scenes was 2.5%, planned comparisons showed that for the Detection task, CD patients had a lower sensitivity than controls (CD: 2.41 ± 1.37 ; controls: 3.50 ± 0.44 ; $F(1,44) = 11.55$, $p < 0.05$) but there was no difference between controls and NoCD patients (controls: 3.50 ± 0.44 ; NoCD: 3.13 ± 0.80 ; $F(1,44) = 1.34$, $p = 0.25$). For the Categorization task, CD patients as well as NoCD patients had a lower sensitivity than controls (CD: 0.59 ± 1.04 ; NoCD: 1.21 ± 1.19 ; controls: 2.21 ± 0.84 ; CD vs. controls: $F(1,44) = 21.28$, $p < 0.001$; NoCD vs. controls: $F(1,44) = 8.15$, $p < 0.01$).

We were interested to see how the level of contrast influenced the difference between detection and categorization performances in each patient group. We first examined the three-way interaction between Group, Task and Contrast, considering only controls and CD patients. This analysis revealed that Contrast had an altered impact on the difference between detection and categorization in these two groups ($F(1,44) = 5.83$, $p < 0.05$). Indeed, for controls, the difference was larger at 2.5% than at 10% contrast (10% = 0.57 ± 0.73 ; vs. 2.5% = 1.29 ± 0.85 ; Task \times Contrast interaction: $F(1,44) = 11.91$, $p < 0.001$), while for CD patients, the difference remained constant between 2.5% and 10% (10% = 2.01 ± 0.92 ; vs. 2.5% = 1.82 ± 1.00 ; Task \times Contrast interaction: $F(1,44) < 1$). Then, we examined the three-way interaction, considering only controls and NoCD patients. This interaction was not significant ($F(1,44) < 1$), suggesting that the contrast level did not have a greater influence on the difference between detection and categorization for NoCD patients (10% = 0.96 ± 0.85 ; vs. 2.5% = 1.92 ± 0.96) than for controls.

Face experiment

The analysis of variance performed on d' first revealed a main effect of the group ($F(2,44) = 14.88$, $p < 0.001$, $\eta^2 = 0.40$) with a higher sensitivity for controls than for CD patients (controls: 3.73 ± 0.23 ; CD: 2.87 ± 0.73 ; $F(1,44) = 28.71$, $p < 0.001$). However, NoCD patients performed well overall. There was no difference in sensitivity between controls and NoCD patients (controls: 3.73 ± 0.23 ; NoCD: 3.54 ± 0.42 ; $F(1,44) = 2.96$, $p = 0.09$). The main effect of the Task was significant ($F(2,44) = 24.43$, $p < 0.001$, $\eta^2 = 0.36$), with a lower sensitivity for the Categorization task (3.30 ± 0.76) than for the Detection task (3.67 ± 0.45). Furthermore, the Task interacted with the Group factor (Fig 4B; $F(2,44) = 5.31$, $p < 0.05$, $\eta^2 = 0.19$). Planned comparisons revealed that, for the Detection task, CD patients had a lower sensitivity than controls (CD: 3.29 ± 0.65 ; controls: 3.86 ± 0.19 ; $F(1,44) = 16.24$, $F(1,44) = 16.24$, $p < 0.001$), while the difference between NoCD patients and controls was not significant (NoCD: 3.63 ± 0.41 ; controls:

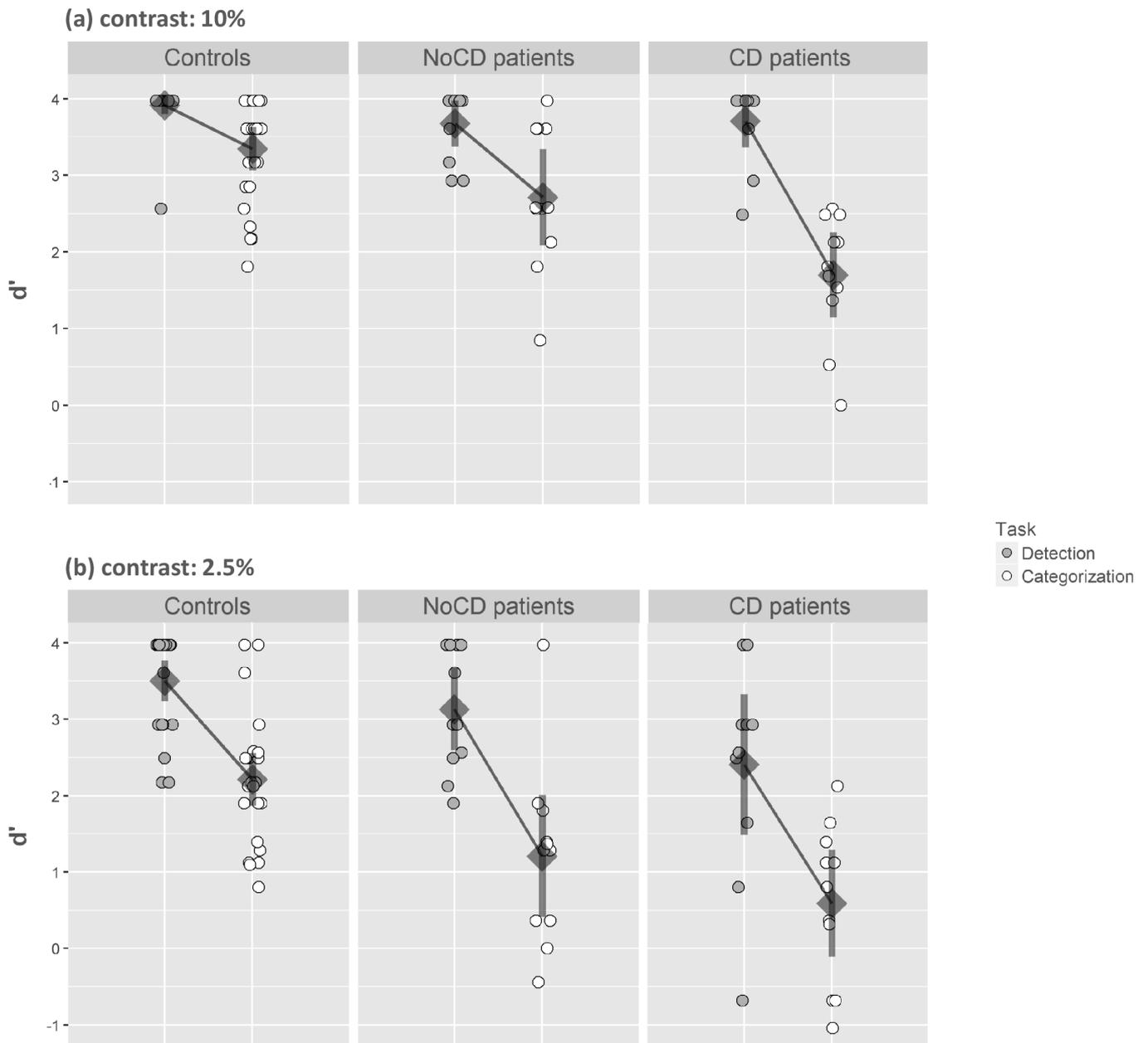


Fig 5. Index of sensitivity (d') in the Scene experiment. Results are presented as a function of the group (primary open-angle glaucoma patients with central defect: CD; primary open-angle glaucoma patients without central defect: NoCD; and normally sighted age-matched participants: controls), the task (detection vs. categorization), and the contrast level (10% vs. 2.5%). Error bars correspond to 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0193465.g005>

3.86 ± 0.19 ; $F(1,44) = 2.63$, $p = 0.11$). Similarly, for the Categorization task, mean comparisons showed that CD patients had a lower sensitivity than controls (CD: 2.45 ± 0.95 ; controls: 3.60 ± 0.44 ; $F(1,44) = 26.80$, $p < 0.001$). However, there was no difference between NoCD patients and controls (NoCD: 3.46 ± 0.54 ; controls: 3.60 ± 0.44 ; $F(1,44) < 1$). Considering only controls and CD patients, the Task interacted with the Group ($F(1,44) = 8.62$, $p < 0.05$) suggesting that the difference between sensitivity for the Categorization task and for the Detection task was larger for CD patients than for controls (CD patients: 0.83 ± 0.70 ; controls:

0.26 ± 0.49). In contrast, considering only controls and NoCD patients as groups, there was no Group × Task interaction (NoCD patients: 0.17 ± 0.47; controls: 0.26 ± 0.49; $F(1,44) < 1$).

The main effect of Contrast was significant ($F(2,44) = 11.83, p < 0.001$) with a lower sensitivity for 2.5% (3.21 ± 1.01) than 10% (3.76 ± 0.41) and this factor interacted with the Task and the Group (Fig 6; $F(2,44) = 4.64, p < 0.05, \eta^2 = 0.17$). When the luminance contrast was 10%, planned comparisons revealed that, for the Detection task, there was no difference in sensitivity between controls and NoCD patients (controls: 3.96 ± 0.07 ; NoCD: 3.97 ± 0.00 ; $F(1,44) < 1$), but controls had a better sensitivity than CD patients (controls: 3.96 ± 0.07 ; CD: 3.71 ± 0.45 ; $F(1,44) = 9.45, p < 0.05$). For the Categorization task, there was no difference between controls and NoCD patients (controls: 3.65 ± 0.49 ; NoCD: 3.87 ± 0.17 ; $F(1,44) = 1.22, p = 0.27$), or between controls and CD patients (controls: 3.65 ± 0.49 ; CD: 3.33 ± 0.88 ; $F(1,44) = 2.45, p = 0.12$). Again, when the luminance contrast was 2.5%, planned comparisons showed that for the Detection task, CD patients had a lower sensitivity than controls (CD: 2.86 ± 1.25 ; controls: 3.76 ± 0.39 ; $F(1,44) = 10.46, p < 0.05$) but there was no difference between controls and NoCD patients (NoCD: 3.29 ± 0.81 ; controls: 3.76 ± 0.39 ; $F(1,44) = 2.94, p = 0.09$). For the Categorization task, CD patients had a lower sensitivity than controls (CD: 1.58 ± 1.25 ; controls: 3.55 ± 0.60 ; $F(1,44) = 38.56, p < 0.001$). In contrast, there was no difference between NoCD patients and controls (NoCD: 3.05 ± 0.99 ; controls: 3.55 ± 0.60 ; $F(1,44) = 2.55, p = 0.12$).

We were interested to see how the level of contrast influenced the difference between detection and categorization performances in each patient group, in comparison to controls. We first examined the three-way interaction between Group, Task and Contrast considering only controls and CD patients. This analysis revealed that Contrast had a different impact on the categorization effect in these two groups ($F(1,44) = 9.24, p < 0.05$). For controls, the difference between detection and categorization did not change between 2.5% and 10% contrast (10% = 0.31 ± 0.50 ; vs. 2.5% = 0.21 ± 0.75 ; $F(1,44) < 1$), while for CD patients, the difference between detection and categorization was larger at 2.5% than at 10% contrast (10% = 0.38 ± 1.10 ; vs. 2.5% = 1.29 ± 0.75 ; $F(1,44) = 10.76, p < 0.05$). Then, we examined the three-way interaction considering only controls and NoCD patients. Once again, this interaction was not significant ($F(1,44) < 1$; NoCD patients: 10% = 0.10 ± 0.17 ; vs. 2.5% = 0.24 ± 0.81).

Discussion

This study was aimed at investigating, in the central visual field, visual abilities in the detection and categorization of photographs of scenes and faces of patients suffering from POAG. The Results did not show the same outcomes for patients, depending on the presence of a central visual defect in static perimetry evaluation using Humphrey 24–2 SITA-Standard test. In comparison with age-matched controls, CD patients showed a deficit in both detection and categorization of low-contrast scene and face stimuli. Furthermore, the difference in performance between detection and categorization was greater for CD patients than for the controls, suggesting that the cost of high-level visual processes for patients was higher. It should be noted that, for the face categorization task, the results of CD patients for face stimuli at 2.5% contrast support those from other behavioral studies [6,11]. These studies showed that patients who were at an advanced stage of the disease exhibit a deficit in a face recognition task, but that patients could adapt to this deficit using specific eye movements patterns. However, in the present study, stimuli were presented for 100 ms, thus making it impossible to rely on eye movements strategies. Unexpectedly, for the categorization of face stimuli at 10% contrast, we did not observe significant difference between controls and CD patients. This result can be explained by the fact that the retinal sensitivity defect in CD patients does not cover the entire central 6° region, in which face stimuli were presented. Several studies on face perception have

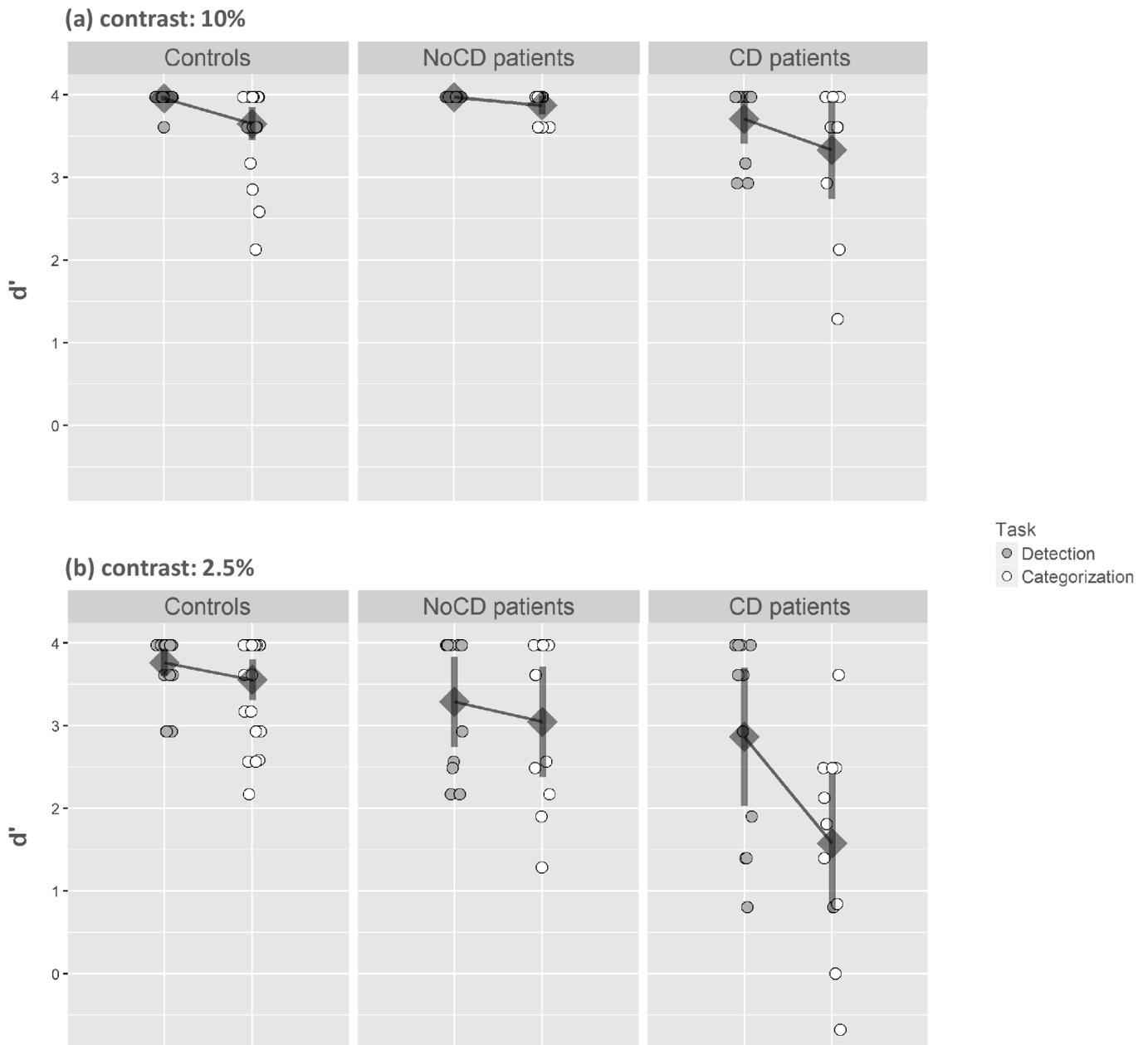


Fig 6. Index of sensitivity (d') in the Face experiment. Results are presented as a function of the group (primary open-angle glaucoma patients with central defect: CD; primary open-angle glaucoma patients without central defect: NoCD; and normally sighted age-matched participants: controls), the task (detection vs. categorization), and the contrast level (10% vs. 2.5%). Error bars correspond to 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0193465.g006>

shown that face processing can be based on featural processing [21,22,23,24]. Thus, CD patients, even with their partial representation of the image, may rely on face features to perform a supraordinate categorization of human faces among other stimuli (animals and vehicles images). The partial, and not uniform, perimetry defect in the central visual field of CD patients could thus be strongly detrimental to the global perception of a stimulus displayed in the whole central visual field. Therefore, the study showed that CD patients have a deficit in

categorization that surpasses their compromised detection. This suggests that visual tests based on stimuli detection may be not adequate to estimate the extent of visual disabilities of patients in their daily life. NoCD patients showed similar performance to age-matched controls for the detection and categorization of faces. For scenes, patients' performances did not significantly differ from those of controls for the detection task, but they demonstrated a deficient categorization task. These results point to a high-level visual deficit that is not predicted by simple retinal sensitivity in the central visual field of patients. Patients participated in this experiment following their usual ophthalmological examination in which a central visual field defect was identified using a 24–2 SITA-Standard procedure of the Humphrey visual field analyzer. In clinical practice, if this procedure does not show a central visual field defect, a complementary 10–2 SITA-procedure is not performed. The latter could have resulted in a better detection of an early central defect [25].

A major conclusion of our results is that they suggest that a peripheral vision loss may induce a subtle high-level visual deficit in central vision. In spite of its low resolution, peripheral vision in normally sighted people is quite efficient and allows the recognition of scenes and objects, even at very high eccentricities (above 70°) [26, 27]. Furthermore, peripheral information could influence the recognition of information by central vision. Indeed, Boucart et al. [26] showed that the categorization of an object as an animal by central vision is facilitated when the animal is congruent to its scene background (relative to an incongruent scene background) in normally sighted aged people, but also in patients with aged-related macular degeneration (i.e. a disease characterized by loss of central vision).

The discrepancy in findings between the Scene and Face experiments for NoCD patients can be explained by many factors. First, the Scene experiment requires a subordinate categorization (indoors vs. outdoor scenes), while the Face experiment needs a categorization at the basic level (human faces vs. animals vs. vehicles). According to the cognitive economy principle, the basic level would be favored because it maximizes both similarity within a category (e.g. human faces are structurally very similar), and distinction between categories (e.g. human faces, vehicles and animals are structurally different). Thus, basic-level categorization would be the earliest and the more efficient categorization. A subordinate level categorization would be more costly, involving more cognitive and visual processes [28]. In our experiment, indoor and outdoor scenes have common physical characteristics, in particular similar visual cluttering and distribution of energy in terms of spatial frequency and dominant orientations. Their discrimination may thus rely on an additional process such as identification of details. An alternative explanation would be linked to real-life characteristics of these two categories. Behavioral and neuroimaging studies have shown that the periphery is more useful than central vision for categorizing natural scenes [29,30,31,32,33]. For example, young healthy subjects can categorize a scene as a beach, a forest or a street more accurately when only peripheral information is available (i.e. when the central portion of the scene is hidden by a circular scotoma) than when only the central portion of the scene is available [30]. In natural conditions, global scene perception is constant, encompassing the whole visual field, even when central vision is dedicated to a fine, object-directed visual task. Scenes can thus be considered as a category mostly calling upon peripheral vision, which is also suggested by an fMRI study [31] showing an overlap between regions activated by images of buildings and the cortical representation of the peripheral visual field. In contrast, the same study showed that faces preferentially activated the cortical representation of the central visual field. This may explain why NoCD patients were mostly impaired in the scene experiment, while face categorization performance was relatively highly preserved. It should be noted that our results for the face experiment support those from another behavioral study [6] showing that only POAG patients who were at an advanced stage of the disease exhibit a deficit in a face recognition task.

Based on the results of Lenoble et al. [7], we predicted that the deficit in categorization relative to detection would be stronger with a lower contrast level of stimuli. This hypothesis is also driven by the fact that patients often complain about extreme lighting conditions, like a darkened room, twilight, full sunlight or excessive artificial lighting [4,34]. From a neurophysiological point of view, parasol ganglion cells have higher contrast gain than midget ganglion cells. They are thus more sensitive to low luminance contrasts. These cells form the magnocellular pathway and are found mainly in the peripheral retina [35]. Glaucoma affects the parasol ganglion cells early in the disease. It could thus affect the magnocellular pathway and consequently cause a higher recognition deficit with low contrasts. However, results of the Scene experiment did not show a contrast effect. For NoCD patients, the difference between detection and categorization increased with a lower contrast (d' difference increased from 0.96 for 10% to 1.92 for 2.5%). However, this was also the case for the controls (d' difference increased from 0.57 for 10% to 1.29 for 2.5%). Although the effect of contrast was greater for patients, the statistical analysis did not show a significant interaction. For CD patients, the difference between detection and categorization was not statistically different between 10% and 2.5% and sensitivity showed a floor effect at 2.5%. Therefore, our study revealed a contrast effect on high-level processes involved in scene recognition in both aged controls and NoCD patients. Nevertheless it did not detect a greater contrast effect in NoCD patients than in the controls. It should be noted that the contrast levels used by Lenoble et al. [7] were 50% and 100%, i.e. higher and more distinctive than those used in the present study (2.5% and 10%). These very low contrast levels were initially chosen to simulate twilight conditions. They may have been too low. Hence they could have significantly biased effects. Further studies need to increase the number of contrast levels to establish the critical contrast range for which visual recognition difficulties in POAG patients are most likely to be measured.

Importantly, results of the present study showed that compared to age-matched controls, NoCD patients, demonstrated a deficit for scene categorization in their central vision. The detection of scenes was well-maintained. Thus, the deficit for categorization in central vision can hardly be explained by the sole retinal sensitivity loss for these patients. An alternative explanation might be that the progressive destruction of retinal ganglion cells in glaucoma can trigger trans-synaptic degeneration in the lateral geniculate nucleus and visual cortex [36,37,38]. This would cause structural and functional changes in high-level cerebral areas, affecting visual function as a whole. Increasing evidence from MRI studies supports the hypothesis of important anatomical and functional cortical changes linked to a neuronal degeneration in glaucoma [39,40,41,42,43,44,45]. For example, a voxel-based morphometry study [39] revealed that grey matter density of patients with POAG was reduced in the medial part of the anterior occipital cortex compared to controls. This corresponds to the approximate visual field defect projections of patients in the visual cortex. Studies using functional MRI retinotopic mapping allowed exploration of activation of the primary visual cortex, with respect to the representation of the visual field in patients with POAG. They revealed an alteration of the MRI signal consistent with the peripheral visual field loss as stated by perimetry [40]. In addition to cortical differences directly linked to retinal loss, other studies found abnormalities in many other visual and non-visual brain areas of POAG patients [41,44,45,46,47,34,35,36,37]. For example, a high resolution MRI study [44] showed that the optic nerve and the optic chiasm were thinner in patients than in controls. This study also showed that the grey matter density was reduced in the calcarine and lingual gyri. It was also decreased or increased in numerous temporal, frontal, and parietal cortical regions. Similarly, another MRI study [41] showed a grey matter atrophy in cortical regions involved in object and scene recognition (i.e. the lateral occipital complex or LOC [48] and the parahippocampal place area or PPA, respectively [17,49,50]). The study revealed a decrease in functional

connectivity within the visual cortical network but also within the working memory and the attention networks. Concerning specifically the visual cortical network, a resting state fMRI study [47] reported alteration of functional connectivity between early visual areas and high-order visual areas dedicated to visual recognition like the fusiform and the inferior temporal gyri. Overall, MRI studies of glaucoma point out evidence for a large range of structural and functional modifications in both the visual and non-visual brain. In particular, alterations in integrative visual areas supporting the final step of visual recognition in normal brain could explain why patients with or without central visual defects may suffer from scene recognition impairment. Therefore, studying functional and structural abnormalities of integrative high-level visual areas (e.g. the scene selective PPA) in relation to behavioral measures seems to be an interesting approach for the understanding of visual scene recognition deficits in POAG patients.

In conclusion, this study highlights an alteration of visual recognition in the central visual field of patients with glaucoma, despite the absence of a central defect in static perimetry evaluation using Humphrey 24–2 SITA-Standard test. Our experimental paradigm compares the abilities of POAG patients to detect and categorize the same stimuli, allowing support of the hypothesis of a deficit in categorization even when detection abilities are intact.

Acknowledgments

We thank Katherine M. Kean, a native English-speaking life scientist, and Jean-François Nominé, both of the Translation Unit in the Institute for scientific and technical information-CNRS for the revision (<http://www.inist.fr/>) of the English of the manuscript.

Author Contributions

Conceptualization: Alexia Roux-Sibilon, Floriane Rutgé, Arnaud Attye, Nathalie Guyader, Muriel Boucart, Carole Peyrin.

Data curation: Floriane Rutgé.

Formal analysis: Alexia Roux-Sibilon, Carole Peyrin.

Funding acquisition: Floriane Rutgé, Florent Aptel, Carole Peyrin.

Investigation: Alexia Roux-Sibilon, Carole Peyrin.

Methodology: Alexia Roux-Sibilon, Floriane Rutgé, Nathalie Guyader, Carole Peyrin.

Resources: Alexia Roux-Sibilon, Florent Aptel, Nathalie Guyader.

Supervision: Florent Aptel, Christophe Chiquet, Carole Peyrin.

Validation: Alexia Roux-Sibilon, Floriane Rutgé, Florent Aptel, Carole Peyrin.

Visualization: Alexia Roux-Sibilon, Floriane Rutgé.

Writing – original draft: Alexia Roux-Sibilon, Carole Peyrin.

Writing – review & editing: Alexia Roux-Sibilon, Floriane Rutgé, Florent Aptel, Arnaud Attye, Nathalie Guyader, Muriel Boucart, Christophe Chiquet, Carole Peyrin.

References

1. Hood DC, Raza AS, de Moraes CG V, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res.* 2013. <https://doi.org/10.1016/j.preteyeres.2012.08.003> PMID: 22995953

2. Crabb DP, Smith ND, Glen FC, Burton R, Garway-Heath DF. How does glaucoma look?: Patient perception of visual field loss. *Ophthalmology* [Internet]. 2013; 120(6):1120–6. Available from: <http://dx.doi.org/10.1016/j.ophtha.2012.11.043> PMID: 23415421
3. Crabb DP. A view on glaucoma—Are we seeing it clearly? *Eye*. 2016; 30(2):304–13. <https://doi.org/10.1038/eye.2015.244> PMID: 26611846
4. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol*. 2009; 20(2):92–98. <https://doi.org/10.1097/ICU.0b013e32832401a9> PMID: 19240541
5. McKendrick AM, Badcock DR, Morgan WH. The detection of both global motion and global form is disrupted in glaucoma. *Investig Ophthalmol Vis Sci*. 2005. <https://doi.org/10.1167/iovs.04-1406> PMID: 16186351
6. Glen FC, Crabb DP, Smith ND, Burton R, Garway-Heath DF. Do patients with glaucoma have difficulty recognizing faces? *Investig Ophthalmol Vis Sci*. 2012; 53(7):3629–3637. <https://doi.org/10.1167/iovs.11-8538> PMID: 22511628
7. Lenoble Q, Lek JJ, Mckendrick AM. Visual object categorisation in people with glaucoma. <https://doi.org/10.1136/bjophthalmol-2015-308289>
8. Kotecha A, O'Leary N, Melmoth D, Grant S, Crabb DP. The functional consequences of glaucoma for eye-hand coordination. *Investig Ophthalmol Vis Sci*. 2009; 50(1):203–213. <https://doi.org/10.1167/iovs.08-2496> PMID: 18806294
9. Dive S, Rouland JF, Lenoble Q, Szaffarczyk S, McKendrick AM, Boucart M. Impact of Peripheral Field Loss on the Execution of Natural Actions. *J Glaucoma*. 2016; 25(10):e889–e896. <https://doi.org/10.1097/JG.0000000000000402> PMID: 26950583
10. Burton R, Smith ND, Crabb DP. Eye movements and reading in glaucoma: observations on patients with advanced visual field loss. *Graefe's Arch Clin Exp Ophthalmol*. 2014; 252(10):1621–30.
11. Glen FC, Smith ND, Crabb DP. Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Res* [Internet]. 2013; 82:42–51. Available from: <http://dx.doi.org/10.1016/j.visres.2013.02.010> PMID: 23485426
12. Smith ND, Glen FC, Crabb DP. Eye movements during visual search in patients with glaucoma. *BMC Ophthalmol*. 2012; 12(1):45. <https://doi.org/10.1186/1471-2415-12-45> PMID: 22937814
13. Crabb DP, Smith ND, Rauscher FG, Chisholm CM, Barbur JL, Edgar DF, et al. Exploring eye movements in patients with glaucoma when viewing a driving scene. *PLoS One*. 2010; 5(3).
14. Crabb DP, Smith ND, Zhu H. What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. *Front Aging Neurosci*. 2014; 6(NOV):1–10. <https://doi.org/10.3389/fnagi.2014.00001>
15. Smith ND, Crabb DP, Garway-Heath DF. An exploratory study of visual search performance in glaucoma. *Ophthalmic Physiol Opt*. 2011; 31(3):225–32. <https://doi.org/10.1111/j.1475-1313.2011.00836.x> PMID: 21470272
16. Hawkins AS, Szyk JP, Ardickas Z, Alexander KR, Wilensky JT. Comparison of contrast sensitivity, visual acuity, and Humphrey visual field testing in patients with glaucoma. *J Glaucoma*. 2003; 12:134–138. <https://doi.org/10.1097/00061198-200304000-00008> PMID: 12671468
17. Kauffmann L, Ramanoël S, Guyader N, Chauvin A, Peyrin C. Spatial frequency processing in scene-selective cortical regions. *Neuroimage*. 2015. <https://doi.org/10.1016/j.neuroimage.2015.02.058> PMID: 25754068
18. Ramanoël S, Kauffmann L, Cousin E, Dojat M, Peyrin C. Age-related differences in spatial frequency processing during scene categorization. *PLoS One*. 2015; 10(8). <https://doi.org/10.1371/journal.pone.0134554> PMID: 26288146
19. Guyader N, Chauvin A, Boucart M, Peyrin C. Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision Res*. 2017; 133:100–111. <https://doi.org/10.1016/j.visres.2016.12.019> PMID: 28202396
20. Bex PJ, Makous W. Spatial frequency, phase, and the contrast of natural images. *J Opt Soc Am A Opt Image Sci Vis*. 2002; 19(6):1096–1106. <https://doi.org/10.1364/JOSAA.19.001096> PMID: 12049346
21. Goffaux V, Rossion B. Face Inversion Disproportionately Impairs the Perception of Vertical but not Horizontal Relations Between Features. *J Exp Psychol Hum Percept Perform*. 2007; 33(4):995–1002. <https://doi.org/10.1037/0096-1523.33.4.995> PMID: 17683242
22. Mckone E, Yovel G. Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? toward a new theory of holistic processing. *Psychon Bull Rev*. 2009; 16(5):778–97. <https://doi.org/10.3758/PBR.16.5.778> PMID: 19815781
23. Tanaka JW, Kaiser MD, Hagen S, Pierce LJ. Losing face: Impaired discrimination of featural and configural information in the mouth region of an inverted face. *Attention, Perception, Psychophys*. 2014; 76(4):1000–14.

24. Cabeza R, Kato T. Features are Also Important: Contributions of Featural and Configural Processing to Face Recognition. *Psychol Sci* [Internet]. 2000; 11(5):429–33. Available from: <http://journals.sagepub.com/doi/10.1111/1467-9280.00283> PMID: 11228917
25. Marechal M, Chaprin T, El Chehab H, Fenolland JR, Marill AF, Rosenberg R, et al. Interest of central 10-2 visual field in glaucomatous and glaucoma suspects patients with normal central 24-2 visual field. *Acta Ophthalmol*. 92(s253):0–0. <https://doi.org/10.1111/j.1755-3768.2014.T039.x>
26. Boucart M, Moroni C, Szaffarczyk S, Tran THC. Implicit processing of scene context in macular degeneration. *Investig Ophthalmol Vis Sci*. 2013. <https://doi.org/10.1167/iovs.12-9680> PMID: 23425698
27. Boucart M, Lenoble Q, Quettelart J, Szaffarczyk S, Desprez P, Thorpe SJ. Finding faces, animals, and vehicles in far peripheral vision. *J Vis*. 2016; 16:1–13. <https://doi.org/10.1167/16.2.10> PMID: 27404483
28. Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P. Basic objects in natural categories. *Cogn Psychol*. 1976; 8(3):382–439. [https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X)
29. Jebara N, Pins D, Desprez P, Boucart M. Face or building superiority in peripheral vision reversed by task requirements. *Adv Cogn Psychol*. 2009; 5(1):42–53. <https://doi.org/10.2478/v10053-008-0065-5> PMID: 20523849
30. Larson AM, Loschky LC. The contributions of central versus peripheral vision to scene gist recognition. 2009. <https://doi.org/10.1167/9.10.6>
31. Levy I, Hasson U, Avidan G, Hendler T, Malach R. Center–periphery organization of human object areas. *Nat Neurosci*. 2001; 4(5):533–539. <https://doi.org/10.1038/87490> PMID: 11319563
32. Hasson U, Levy I, Behrmann M, Hendler T, Malach R. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*. 2002; 34(3):479–490. [https://doi.org/10.1016/S0896-6273\(02\)00662-1](https://doi.org/10.1016/S0896-6273(02)00662-1) PMID: 11988177
33. Malach R, Levy I, Hasson U. The topography of high-order human object areas. *Trends Cogn Sci*. 2002; 6(4):176–184. [https://doi.org/10.1016/S1364-6613\(02\)01870-3](https://doi.org/10.1016/S1364-6613(02)01870-3) PMID: 11912041
34. Nelson P, Aspinall P, O'Brien C. Patients' perception of visual impairment in glaucoma: a pilot study. *Br J Ophthalmol*. 1999; 83(5):546–552. <https://doi.org/10.1136/bjo.83.5.546> PMID: 10216052
35. Dacey D. Origins of perception: retinal ganglion cell diversity and the creation of parallel visual pathways. In: *The Cognitive Neurosciences*, 3.; 2004:281–301.
36. Yücel YH, Zhang Q, Weinreb RN, Kaufman PL, Gupta N. Atrophy of relay neurons in magno- and parvocellular layers in the lateral geniculate nucleus in experimental glaucoma. *Investig Ophthalmol Vis Sci*. 2001; 42(13):3216–3222. <https://doi.org/10.1001/archophth.118.3.378>
37. Weber AJ, Chen H, Hubbard WC, Kaufman PL. Number in the Primate Lateral Geniculate Nucleus. *Invest Ophthalmol Vis Sci*. 2000; 41(6):1370–1379. PMID: 10798652
38. Gupta N, Yücel YH. What Changes Can We Expect in the Brain of Glaucoma Patients? *Surv Ophthalmol*. 2007; 52(6 SUPPL.):122–126. <https://doi.org/10.1016/j.survophthal.2007.08.006> PMID: 17998036
39. Boucard CC, Hernowo AT, Maguire RP, Jansonius NM, Roerdink JB, Hooymans JM, et al. Changes in cortical grey matter density associated with long-standing retinal visual field defects. *Brain*. 2009; 132(Pt 7):1898–1906. <https://doi.org/10.1093/brain/awp119> PMID: 19467992
40. Duncan RO, Sample PA, Weinreb RN, Bowd C, Zangwill LM. Retinotopic organization of primary visual cortex in glaucoma: Comparing fMRI measurements of cortical function with visual field loss. *Prog Retin Eye Res*. 2007; 26(1):38–56. <https://doi.org/10.1016/j.preteyeres.2006.10.001> PMID: 17126063
41. Frezzotti P, Giorgio A, Motolese I, De Leucio A, Iester A, Motolese E, et al. Structural and functional brain changes beyond visual system in patients with advanced glaucoma. *PLoS One*. 2014; 9(8). <https://doi.org/10.1371/journal.pone.0105931> PMID: 25162716
42. Gerente VM, Schor RR, Chaim KT, de Maria Felix M, Ventura DF, Teixeira SH, et al. Evaluation of glaucomatous damage via functional magnetic resonance imaging, and correlations thereof with anatomical and psychophysical ocular findings. *PLoS One*. 2015; 10(5):1–12. <https://doi.org/10.1371/journal.pone.0126362> PMID: 25969982
43. Qing G, Zhang S, Wang B, Wang N. Functional MRI Signal Changes in Primary Visual Cortex Corresponding to the Central Normal Visual Field of Patients with Primary Open-Angle Glaucoma. *Investig Ophthalmology Vis Sci*. 2010; 51(9):4627. <https://doi.org/10.1167/iovs.09-4834> PMID: 20357191
44. Chen WW, Wang N, Cai S, Fang Z, Yu M, Wu Q, et al. Structural brain abnormalities in patients with primary open-angle glaucoma: A study with 3T MR imaging. *Investig Ophthalmol Vis Sci*. 2013; 54(1):545–554. <https://doi.org/10.1167/iovs.12-9893> PMID: 23258150
45. Li C, Cai P, Shi L, Lin Y, Zhang J, Suqing L, et al. Voxel-based Morphometry of the Visual-related Cortex in Primary Open Angle Glaucoma. *Curr Eye Res*. 2012; 37(9):794–802. <https://doi.org/10.3109/02713683.2012.683506> PMID: 22631870

46. Brown HDH, Woodall RL, Kitching RE, Baseler HA, Morland AB. Using magnetic resonance imaging to assess visual deficits: A review. *Ophthalmic Physiol Opt.* 2016; 36(3):240–265. <https://doi.org/10.1111/opo.12293> PMID: 27112223
47. Dai H, Morelli JN, Ai F, Yin D, Hu C, Xu D, et al. Resting-state functional MRI: Functional connectivity analysis of the visual cortex in primary open-angle glaucoma patients. *Hum Brain Mapp.* 2013; 34(10):2455–2463. <https://doi.org/10.1002/hbm.22079> PMID: 22461380
48. Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. In: *Vision Research.*; 2001. [https://doi.org/10.1016/S0042-6989\(01\)00073-6](https://doi.org/10.1016/S0042-6989(01)00073-6)
49. Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature.* 1998; 392(6676):598–601. <https://doi.org/10.1038/33402> PMID: 9560155
50. Epstein R. The cortical basis of visual scene processing. *Vis cogn.* 2005; 12(6):954–978. <https://doi.org/10.1080/13506280444000607>

Annexe 4

Annexe 4 - Tableau 1. Coordonnées des pics d'activation des régions impliquées dans la catégorisation de scènes dans l'Expérience 7, pour chaque participant contrôle et chaque patient atteint de glaucome. Nous avons retenu les clusters d'activation sur la base de l'intensité de la réponse individuelle des voxels (à un seuil de 0.001) et de l'étendue spatiale de l'activation (clusters d'au moins 10 voxels).

[Catégorisation > Repos]							
	Contrôle 1	Contrôle 2	Contrôle 3	Contrôle 4	Glaucome 1	Glaucome 2	Glaucome 3
Cortex occipital							
Cuneus, gyrus lingual, scissure calcarine		[-9x -94y 2z]		[-15x -79y -13z] [18x -85y -10z]	[12x -94y -7z]	[-18x -82y -16z]	[-6x -97y -4z]
Gyrus occipital moyen	[24x -94y 11z] [-27x -94y 20z]	[33x -73y 26z] [-27x -70y 29z]	[33x -88y 2z] [-30x -91y 17z]	[36x -76y 11z] [-30x -85y 17z]	[33x -79y 11z] [-21x -91y 8z]	[30x -82y 14z] [-30x -88y 2z]	[39x -76y 23z] [-27x -67y 35z]
Gyrus occipital supérieur			[30x -73y 32z] [-27x -76y 29z]		[27x -70y 35z] [-21x -88y 29z]		
Cortex occipito-temporal							
Gyrus parahippocampique		[30x -46y -10z] [-21x -46y -10z]	[36x -40y -13z] [-33x -46y -4z]	[33x -43y -22z] [-30x -46y -22z]	[33x -37y -13z] [-33x -49y -13z]		
Gyrus fusiforme	[-39x -58y -16z]					[-39x -76y -16z]	
Cortex frontal et insulaire							
Insula	[-33x -23y -5z]						[48x 14y -4z] [-33x 20y -4z]
Gyrus frontal inférieur (<i>pars orb.</i>)		[27x 23y -22z]	[-51x 29y -7z]	[45x 44y -7z] [-42x 41y -7z]			[54x 32y -7z] [-51x 41y 2z]
Gyrus frontal inférieur (<i>pars tri.</i>)	[-48x 29y 17z]	[39x 32y 14z] [-45x 20y 23z]	[48x 35y 26z] [-42x 17y 26z]	[48x 38y 23z] [-45x 32y 23z]	[48x 23y 20z]		[45x 38y 23z] [-39x 29y 26z]
Gyrus frontal inférieur (<i>pars oper.</i>)	[42x 8y 32z] [-39x 5y 35z]			[54x 14y -1z] [39x 2y 26z]			

Annexe 4 - Tableau 2. Pour le niveau de contraste RMS = 0.03, coordonnées des pics d'activation des régions impliquées dans la catégorisation de scènes dans l'Expérience 7, pour chaque participant contrôle et chaque patient atteint de glaucome. Nous avons retenu les clusters d'activation sur la base de l'intensité de la réponse individuelle des voxels (à un seuil de 0.001) et de l'étendue spatiale de l'activation (clusters d'au moins 10 voxels).

[Catégorisation > Repos] (RMS=0.03)							
	Contrôle 1	Contrôle 2	Contrôle 3	Contrôle 4	Glaucome 1	Glaucome 2	Glaucome 3
Cortex occipital							
Cuneus, gyrus lingual, scissure calcarine		[9x -97y -7z]					[-6x -97y -4z]
Gyrus occipital moyen	[-27x -85y 29z]	[-12x -100y 2z]	[-33x -88y 2z] [27x -91y 2z]	[-27x -82y 17z] [30x -73y 20z]		[30x -85y 11z]	[36x -73y 23z]
Gyrus occipital supérieur	[24x -94y 11z]			[21x -94y 14z]	[27x -82y 23z] [-21x -88y 29z]		
Cortex occipito-temporal							
Gyrus parahippocampique		[-21x -46y -10z] [30x -46y -10z]		[30x -40y -22z] [-30x -46y -22z]	[33x -37y -13z]		
Gyrus fusiforme	[-39x -40y -19z]	[42x -49y -22z]	[48x -46y -22z] [-39x -46y -25z]	[30x -40y -22z] [-30x -46y -22z]	[33x -49y -13z]		
Cortex frontal et insulaire							
Insula	[-33x -23y 8z]	[-27x -26y 5z] [-30x 23y -4z]			[42x 17y -1z]		[48x 14y -7z] [-33x 20y -4z]
Gyrus frontal inférieur (<i>pars orb.</i>)			[-48x 26y -10z]	[-42x 41y -7z]	[45x 29y -4z]		[54x 32y -7z]
Gyrus frontal inférieur (<i>pars tri.</i>)	[48x -29y 20z]	[42x 32y 14z] [-48x 20y 23z]	[48x 35y 26z]	[45x 38y 23z]	[51x 23y 23z] [-45x 14y 23z]		[-48x 44y 8z]
Gyrus frontal inférieur (<i>pars oper.</i>)	[42x 8y 32z]	[45x 17y 26z]	[-51x 20y 32z]				[-39x 8y 26z]

Annexe 4 - Tableau 3. Pour le niveau de contraste RMS = 0.11, coordonnées des pics d'activation des régions impliquées dans la catégorisation de scènes dans l'Expérience 7, pour chaque participant contrôle et chaque patient atteint de glaucome. Nous avons retenu les clusters d'activation sur la base de l'intensité de la réponse individuelle des voxels (à un seuil de 0.001) et de l'étendue spatiale de l'activation (clusters d'au moins 10 voxels).

[Catégorisation > Repos] (RMS=0.11)							
	Contrôle 1	Contrôle 2	Contrôle 3	Contrôle 4	Glaucome 1	Glaucome 2	Glaucome 3
Cortex occipital							
Cuneus, gyrus lingual, scissure calcarine		[-9x -94y -2z]	[-15x -64y 20z]	[18x -85y -10z]		[-18x -82y -16z]	
Gyrus occipital moyen	39x -85y 5z [-27x -85y 29z]	[30x -70y -26z]	[-30x -91y 17z] [30x -88y 5z]	[-27x -91y 8z]		[33x -91y 8z]	[36x -82y 23z] [-27x -64y 38z]
Gyrus occipital supérieur	[24x -94y 11z]						
Cortex occipito-temporal							
Gyrus parahippocampique		[-21x -46y -10z] [30x -49y -10z]		[-30x -46y -22z]			
Gyrus fusiforme	[-39x -61y -16z]	[42x -49y -22z]	[30x -55y -4z]	[33x -40y -22z] [-30x -46y -22z]		[-36x -49y -16z]	[-39x -76y -16z]
Cortex frontal et insulaire							
Insula	[33x -23y 5z]						[42x 17y 2z]
Gyrus frontal inférieur (<i>pars orb.</i>)		[24x 17y -19z]		[45x 41y -7z] [-39x 41y -4z]			[51x 35y -7z] [-39x 5y 26z]
Gyrus frontal inférieur (<i>pars tri.</i>)	[48x -29y 17z] [-51x -47y 11z]			[48x 38y 23z] [-45x 32y 23z]			[39x 20y 26z] [-48x 44y 5z]
Gyrus frontal inférieur (<i>pars oper.</i>)	[42x 8y 32z]		[-51x 17y 32z]	[-54x 14y -1z]			[-51x 11y -1z]

Bibliographie

- Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative neurology and neuroscience*, *26*(4, 5), 435–446.
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.
- Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of vision*, *12*(10), 18–18.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122–131.
- Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of neuroscience*, *29*(34), 10638–10652.
- Ariely, D. (2001). Seeing sets : Representation by statistical properties. *Psychological science*, *12*(2), 157–162.
- Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., & Kouider, S. (2013). Nonconscious learning from crowded sequences. *Psychological science*, *25*(1), 113–119.
- Azzopardi, P., Jones, K. E., & Cowey, A. (1999). Uneven mapping of magnocellular and parvocellular projections from the lateral geniculate nucleus to the striate cortex in the macaque monkey. *Vision research*, *39*(13), 2179–2189.
- Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing : Backward masking and natural scene categorisation. *Vision research*, *45*(11), 1459–1469.
- Badcock, J. C., Whitworth, F. A., Badcock, D. R., & Lovegrove, W. J. (1990). Low-frequency filtering and the processing of local—global stimuli. *Perception*, *19*(5), 617–629.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of vision*, *9*(12), 13–13.
- Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in

- neural scene-processing networks during natural viewing. *Journal of vision*, 16(2), 9–9.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*, 15(4), 600–609.
- Bar, M. (2007). The proactive brain : using analogies and associations to generate predictions. *Trends in cognitive sciences*, 11(7), 280–289.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... others (2006). Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*, 103(2), 449–454.
- Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, 25(3), 343–352.
- Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4 : Linear mixed-effects models using eigen and s4. *R package version*, 1(7), 1–23.
- Bauer, B. (2009). Does stevens’s power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, 59(2), 171–185.
- Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior : integrating self-monitoring and emotion-cognition interactions. *Journal of cognitive neuroscience*, 18(6), 871–879.
- Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *JOSA A*, 19(6), 1096–1106.
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177(4043), 77–80.
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of physiology*, 203(1), 237–260.
- Bonnar, L., Gosselin, F., & Schyns, P. G. (2002). Understanding dali’s slave market with the disappearing bust of voltaire : A case study in the scale information driving perception. *Perception*, 31(6), 683–691.
- Booth, M., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral cortex (New York, NY : 1991)*, 8(6), 510–523.
- Bossaerts, P. (2010). Risk and risk prediction error signals in anterior insula. *Brain Structure and Function*, 214(5-6), 645–653.
- Boucard, C. C., Hernowo, A. T., Maguire, R. P., Jansonius, N. M., Roerdink, J. B., Hooymans, J. M., & Cornelissen, F. W. (2009). Changes in cortical grey matter density associated with long-standing retinal visual field defects. *Brain*, 132(7), 1898–1906.
- Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Desprez, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of vision*, 16(2), 10–10.
- Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, 86, 35–42.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241),

- 177.
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fmri and meg decoding. *Journal of Neuroscience*, *37*(32), 7700–7710.
- Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision research*.
- Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature neuroscience*, *7*(8), 880.
- Brown, H. D., Woodall, R. L., Kitching, R. E., Baseler, H. A., & Morland, A. B. (2016). Using magnetic resonance imaging to assess visual deficits : a review. *Ophthalmic and Physiological Optics*, *36*(3), 240–265.
- Brown, R. (1958). Words and things.
- Bullier, J. (2001). Integrated model of visual processing. *Brain research reviews*, *36*(2-3), 96–107.
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision : the first 1000 ms. *Journal of vision*, *13*(10), 1–1.
- Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. a review. *Cerebral cortex*, *10*(3), 220–242.
- Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2013). Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral cortex*, *24*(11), 2899–2907.
- Chen, W. W., Wang, N., Cai, S., Fang, Z., Yu, M., Wu, Q., . . . others (2013). Structural brain abnormalities in patients with primary open-angle glaucoma : a study with 3t mr imaging. *Investigative ophthalmology & visual science*, *54*(1), 545–554.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393–404.
- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181–204.
- Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of vision*, *14*(8), 5–5.
- Crabb, D. P. (2016). A view on glaucoma—are we seeing it clearly? *Eye*, *30*(2), 304.
- Crabb, D. P., Smith, N. D., Glen, F. C., Burton, R., & Garway-Heath, D. F. (2013). How does glaucoma look? : patient perception of visual field loss. *Ophthalmology*, *120*(6), 1120–1126.
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces : face detection in just 100 ms. *Journal of vision*, *10*(4), 16–16.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of comparative Neurology*, *300*(1), 5–25.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology*, *292*(4), 497–523.
- Dacey, D. (2004). Origins of perception : retinal ganglion cell diversity and the creation

- of parallel visual pathways. *The cognitive neurosciences*, 3, 281–301.
- Dai, H., Morelli, J. N., Ai, F., Yin, D., Hu, C., Xu, D., & Li, Y. (2013). Resting-state functional mri : Functional connectivity analysis of the visual cortex in primary open-angle glaucoma patients. *Human brain mapping*, 34(10), 2455–2463.
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393–401.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564.
- de Fockert, J., & Wolfenstein, C. (2009). Short article : rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, 62(9), 1716–1722.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues : a study in monkeys and humans. *Vision research*, 40(16), 2187–2200.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5), 545–559.
- De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual Review of Psychology*, 31(1), 309–341.
- Dive, S., Rouland, J. F., Lenoble, Q., Szaffarczyk, S., McKendrick, A. M., & Boucart, M. (2016). Impact of peripheral field loss on the execution of natural actions : a study with glaucomatous patients and normally sighted people. *Journal of glaucoma*, 25(10), e889–e896.
- Dubey, S., Bedi, H., Bedi, M., Matah, P., Sahu, J., Mukherjee, S., & Chauhan, L. (2019). Impact of visual impairment on the wellbeing and functional disability of patients with glaucoma in india. *Journal of Current Ophthalmology*.
- Duncan, R. O., Sample, P. A., Weinreb, R. N., Bowd, C., & Zangwill, L. M. (2007). Retinotopic organization of primary visual cortex in glaucoma : Comparing fmri measurements of cortical function with visual field loss. *Progress in retinal and eye research*, 26(1), 38–56.
- Ehinger, K. A., & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, 16(2), 13–13.
- Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, 12(6), 954–978.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598.
- Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli : A horizontal effect. *Vision research*, 43(12), 1329–1335.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience*, 13(2), 171–180.
- Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of mooney faces.

- Journal of vision*, 9(6), 18–18.
- Fechtner, R. D., & Weinreb, R. N. (1994). Mechanisms of optic nerve damage in primary open angle glaucoma. *Survey of ophthalmology*, 39(1), 23–42.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of vision*, 7(1), 10–10.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12), 2379–2394.
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, 16(3), 553–567.
- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation : Effects of social status and speaking. *Cognition*, 117(3), 319–331.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature neuroscience*, 14(9), 1195.
- Frey, S., & Petrides, M. (2000). Orbitofrontal cortex : A key prefrontal region for encoding information. *Proceedings of the National Academy of Sciences*, 97(15), 8723–8727.
- Frezzotti, P., Giorgio, A., Motolese, I., De Leucio, A., Iester, M., Motolese, E., . . . De Stefano, N. (2014). Structural and functional brain changes beyond visual system in patients with advanced glaucoma. *PloS one*, 9(8), e105931.
- Friston, K. J. (2009). The free-energy principle : a rough guide to the brain? *Trends in cognitive sciences*, 13(7), 293–301.
- Friston, K. J. (2010). The free-energy principle : a unified brain theory? *Nature reviews neuroscience*, 11(2), 127.
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4), 1273–1302.
- Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., & Turner, R. (1995). Analysis of fmri time-series revisited. *Neuroimage*, 2(1), 45–53.
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417–458.
- Fuster, J. M., & Bressler, S. L. (2015). Past makes future : role of pfc in prediction. *Journal of cognitive neuroscience*, 27(4), 639–654.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Gerente, V. M., Schor, R. R., Chaim, K. T., de Maria Felix, M., Ventura, D. F., Teixeira, S. H., . . . Paranhos Jr, A. (2015). Evaluation of glaucomatous damage via functional magnetic resonance imaging, and correlations thereof with anatomical and psychophysical ocular findings. *PloS one*, 10(5), e0126362.
- Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of vision*, 18(9), 9–9.

- Ghashghaei, H., Hilgetag, C. C., & Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage*, *34*(3), 905–923.
- Glen, F. C., & Crabb, D. P. (2015). Living with glaucoma : a qualitative study of functional implications and patients' coping behaviours. *BMC ophthalmology*, *15*(1), 128.
- Glen, F. C., Crabb, D. P., Smith, N. D., Burton, R., & Garway-Heath, D. F. (2012). Do patients with glaucoma have difficulty recognizing faces? *Investigative Ophthalmology & Visual Science*, *53*(7), 3629–3637.
- Glen, F. C., Smith, N. D., & Crabb, D. P. (2013). Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Research*, *82*, 42–51.
- Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. *Psychology of learning and motivation*, *43*, 93–126.
- Glovinsky, Y., Quigley, H. A., & Dunkelberger, G. (1991). Retinal ganglion cell loss is size dependent in experimental glaucoma. *Investigative ophthalmology & visual science*, *32*(3), 484–491.
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.
- 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*(4), 1023.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20–25.
- Grainger, J., Tydgate, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology : Human Perception and Performance*, *36*(3), 673.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537.
- Greene, M. R., & Oliva, A. (2009a). The briefest of glances : The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.
- Greene, M. R., & Oliva, A. (2009b). Recognition of natural scenes from global properties : Seeing the forest without representing the trees. *Cognitive psychology*, *58*(2), 137–176.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, *41*(10-11), 1409–1422.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*,

- 21(1), 191–202.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fmri in the human occipital lobe. *Human brain mapping*, 6(4), 316–328.
- Gross, C. G. (2002). Genealogy of the “grandmother cell”. *The Neuroscientist*, 8(5), 512–518.
- Gupta, N., & Yücel, Y. H. (2007). What changes can we expect in the brain of glaucoma patients? *Survey of ophthalmology*, 52(6), S122–S126.
- Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision research*, 133, 100–111.
- Haberman, J., & Whitney, D. (2009). Seeing the mean : ensemble coding for sets of faces. *Journal of Experimental Psychology : Human Perception and Performance*, 35(3), 718.
- Hansen, B. C., & Essock, E. A. (2006). Anisotropic local contrast normalization : The role of stimulus orientation and spatial frequency bandwidths in the oblique and horizontal effect perceptual anisotropies. *Vision research*, 46(26), 4398–4415.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, 34(3), 479–490.
- Hawkins, A. S., Szlyk, J. P., Ardickas, Z., Alexander, K. R., & Wilensky, J. T. (2003). Comparison of contrast sensitivity, visual acuity, and humphrey visual field testing in patients with glaucoma. *Journal of glaucoma*, 12(2), 134–138.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, 50(1), 243–271.
- Hochstein, S., & Ahissar, M. (2002). View from the top : Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804.
- Hodapp, E., Parrish, P., & Anderson, D. (1993). Clinical decisions in glaucoma. *St. Louis, MO : Mosby & Co*, 1–216.
- Hood, D. C., Raza, A. S., de Moraes, C. G. V., Liebmann, J. M., & Ritch, R. (2013). Glaucomatous damage of the macula. *Progress in retinal and eye research*, 32, 1–21.
- Hot, A., Dul, M. W., & Swanson, W. H. (2008). Development and evaluation of a contrast sensitivity perimetry test for patients with glaucoma. *Investigative ophthalmology & visual science*, 49(7), 3049–3057.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, 148(3), 574–591.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215–243.
- Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision*, 13(4), 20–20.
- Jaeger, T. F. (2008). Categorical data analysis : Away from anovas (transformation or

- not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434–446.
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, 8(13), 11–11.
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context : Fast categorization and object interference. *Vision research*, 47(26), 3286–3297.
- Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision* (pp. 2106–2113).
- Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of vision*, 12(13), 16–16.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area : a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11), 4302–4311.
- Kaplan, E. (2004). The m, p, and k pathways of the primate visual system. *The visual neurosciences*, 1, 481–493.
- Katti, H., Peelen, M. V., & Arun, S. (2017). How do targets, nontargets, and scene context influence real-world object detection? *Attention, Perception, & Psychophysics*, 79(7), 2021–2036.
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015c). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of cognitive neuroscience*, 27(12), 2394–2405.
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015a). Rapid scene categorization : Role of spatial frequency order, accumulation mode and luminance contrast. *Vision research*, 107, 49–57.
- Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015b). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. a dynamic causal modeling study. *Brain and cognition*, 99, 46–56.
- Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific reports*, 9(1), 560.
- Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95.
- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, 8, 37.
- Kauffmann, L., Roux-Sibilon, A., Beffara, B., Mermillod, M., Guyader, N., & Peyrin, C. (2017). How does information from low and high spatial frequencies interact during scene categorization? *Visual Cognition*, 25(9-10), 853–867.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm : a critical review. *Psychological bulletin*, 112(1), 24.

- Kimchi, R., & Pirkner, Y. (2015). Multiple level crowding : Crowding at the object parts level and at the object configural level. *Perception*, *44*(11), 1275–1292.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements : Visual processing speed revisited. *Vision research*, *46*(11), 1762–1776.
- Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371–376.
- Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* (pp. 221–244). Springer.
- 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.
- Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more : expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.
- Kotecha, A., O’Leary, N., Melmoth, D., Grant, S., & Crabb, D. P. (2009). The functional consequences of glaucoma for eye–hand coordination. *Investigative ophthalmology & visual science*, *50*(1), 203–213.
- Kouider, S., Berthet, V., & Faivre, N. (2011). Preference is biased by crowded facial expressions. *Psychological Science*, *22*(2), 184–189.
- Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506–1509.
- Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex : evidence from neuroimaging and neuropsychology. *Progress in neurobiology*, *72*(5), 341–372.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package : tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).
- 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.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and cognition*, *65*(2), 145–168.
- Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics*, *54*(6), 773–784.
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6–6.
- Lee, B. L., Gutierrez, P., Gordon, M., Wilson, M. R., Cioffi, G. A., Ritch, R., . . . Mangione, C. M. (1998). The glaucoma symptom scale : a brief index of glaucoma-specific symptoms. *Archives of ophthalmology*, *116*(7), 861–866.
- Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A*, *20*(7), 1434–1448.
- Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early

- visual cortex. *Proceedings of the National Academy of Sciences*, 98(4), 1907–1911.
- Lenoble, Q., Lek, J. J., & McKendrick, A. M. (2016). Visual object categorisation in people with glaucoma. *British Journal of Ophthalmology*, 100(11), 1585–1590.
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision : Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of vision*, 2(2), 3–3.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature neuroscience*, 4(5), 533.
- Li, C., Cai, P., Shi, L., Lin, Y., Zhang, J., Liu, S., ... others (2012). Voxel-based morphometry of the visual-related cortex in primary open angle glaucoma. *Current eye research*, 37(9), 794–802.
- Linares, D., & López-Moliner, J. (2016). quickpsy : An r package to fit psychometric functions for multiple groups. *The R Journal*, 2016, vol. 8, num. 1, p. 122-131.
- Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, 18(4), 513–536.
- Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene-gist recognition with a 180° visual field. *Journal of vision*, 19(5), 15–15.
- Loued-Khenissi, L., Pfeuffer, A., Einhäuser, W., & Preuschoff, K. (2020). Anterior insula reflects surprise in value-based decision-making and perception. *NeuroImage*, 116549.
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding : Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2), 24–24.
- Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations : you spot the animal faster than the bird. *PloS one*, 4(6), e5927.
- Makowski, D. (2018). The psycho package : an efficient and publishing-oriented workflow for psychological science. *J. Open Source Software*, 3(22), 470.
- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in cognitive sciences*, 6(4), 176–184.
- Manassi, M., & Whitney, D. (2018). Multi-level crowding and the paradox of object recognition in clutter. *Current Biology*, 28(3), R127–R133.
- Marendaz, C., Rousset, S., & Charnallet, A. (2003). Reconnaissance des scènes, des objets et des visages. In A. Delorme & M. Fluckiger (Eds.), *Perception et réalité*. Montréal : Gaëtan Morin Editeur.
- McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems : bridging theory and experiment. *Nature Reviews Neuroscience*, 12(7), 415.
- McKendrick, A. M., Badcock, D. R., & Morgan, W. H. (2004). Psychophysical measurement of neural adaptation abnormalities in magnocellular and parvocellular

- pathways in glaucoma. *Investigative ophthalmology & visual science*, 45(6), 1846–1853.
- McKendrick, A. M., Badcock, D. R., & Morgan, W. H. (2005). The detection of both global motion and global form is disrupted in glaucoma. *Investigative Ophthalmology & Visual Science*, 46(10), 3693–3701.
- McKendrick, A. M., Sampson, G. P., Walland, M. J., & Badcock, D. R. (2007). Contrast sensitivity changes due to glaucoma and normal aging : low-spatial-frequency losses in both magnocellular and parvocellular pathways. *Investigative ophthalmology & visual science*, 48(5), 2115–2122.
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, 32(1), 89–115.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision : two cortical pathways. *Trends in neurosciences*, 6, 414–417.
- Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental brain research*, 227(3), 367–377.
- Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). Erp evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, 48(2), 507–517.
- Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing : Replication and extension. *Neuropsychologia*, 56, 447–458.
- Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J.-F., & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of Cognitive Neuroscience*, 25(8), 1315–1331.
- 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.
- Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of cognitive neuroscience*, 26(10), 2287–2297.
- Musselwhite, M., & Jeffreys, D. (1985). The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. *Vision research*, 25(11), 1545–1555.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (mmn) in basic research of central auditory processing : a review. *Clinical neurophysiology*, 118(12), 2544–2590.
- Näätänen, R., & Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound : a review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Navon, D. (1977). Forest before trees : The precedence of global features in visual perception. *Cognitive psychology*, 9(3), 353–383.
- Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution : insights from

- a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, 7(1), 1–17.
- Nelson, P., Aspinall, P., & O'Brien, C. (1999). Patients' perception of visual impairment in glaucoma : a pilot study. *British Journal of Ophthalmology*, 83(5), 546–552.
- Newcombe, F., Ratcliff, G., & Damasio, H. (1987). Dissociable visual and spatial impairments following right posterior cerebral lesions : Clinical, neuropsychological and anatomical evidence. *Neuropsychologia*, 25(1), 149–161.
- Nobre, A., Coull, J., Frith, C., & Mesulam, M. (1999). Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nature neuroscience*, 2(1), 11.
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In *Extrastriate cortex in primates* (pp. 205–241). Springer.
- Nowak, L. G., Munk, M., Girard, P., & Bullier, J. (1995). Visual latencies in areas v1 and v2 of the macaque monkey. *Visual neuroscience*, 12(2), 371–384.
- Oliva, A. (2005). Gist of the scene. In *Neurobiology of attention* (pp. 251–256). Elsevier.
- 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.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene : A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3), 145–175.
- Osterberg, G. (1935). Topography of the layer of the rods and cones in the human retina. *Acta ophthalmol*, 13(6), 1–102.
- Özgen, E., Sowden, P., Schyns, P., & Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition*, 12(6), 925–937.
- Palmer, t. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3, 519–526.
- Parker, D., & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Research*, 27(8), 1279–1283.
- Parker, D., & Salzen, E. (1977). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision research*, 17(10), 1201–1204.
- Parker, D., Salzen, E., & Lishman, J. (1982). Visual-evoked responses elicited by the onset and offset of sinusoidal gratings : latency, waveform, and topographic characteristics. *Investigative Ophthalmology & Visual Science*, 22(5), 675–680.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, 4(7), 739.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area v4. *Journal of neurophysiology*, 82(5), 2490–2502.
- Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, 186, 103–112.

- Peyrin, C. (2018, Apr). *Recor database*. OSF. Consulté sur osf.io/xjqd4 doi: 10.17605/OSF.IO/XJDQ4
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., . . . Vuilleumier, P. (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.
- Poirel, N., Brazo, P., Turbelin, M.-R., Lecardeur, L., Simon, G., Houdé, O., . . . Dollfus, S. (2010). Meaningfulness and global-local processing in schizophrenia. *Neuropsychologia*, *48*(10), 3062–3068.
- Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest : Developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental psychology*, *44*(1), 245.
- Poirel, N., Pineau, A., & Mellet, E. (2006). Implicit identification of irrelevant local objects interacts with global/local processing of hierarchical stimuli. *Acta psychologica*, *122*(3), 321–336.
- Pokorny, J., & Smith, V. C. (1997). Psychophysical signatures associated with magnocellular and parvocellular pathway contrast gain. *JOSA A*, *14*(9), 2477–2486.
- Polyak, S. L. (1941). The retina.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, *40*(1), 49–70.
- Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966.
- Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience*, *28*(11), 2745–2752.
- Qing, G., Zhang, S., Wang, B., & Wang, N. (2010). Functional mri signal changes in primary visual cortex corresponding to the central normal visual field of patients with primary open-angle glaucoma. *Investigative ophthalmology & visual science*, *51*(9), 4627–4634.
- Qiu, A., Rosenau, B. J., Greenberg, A. S., Hurdal, M. K., Barta, P., Yantis, S., & Miller, M. I. (2006). Estimating linear cortical magnification in human primary visual cortex via dynamic programming. *Neuroimage*, *31*(1), 125–138.
- Quigley, H. A., & Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, *90*(3), 262–267.
- Quigley, H. A., Sanchez, R. M., Dunkelberger, G. R., L'Hernault, N. L., & Baginski, T. A. (1987). Chronic glaucoma selectively damages large optic nerve fibers. *Investigative Ophthalmology & Visual Science*, *28*(6), 913–920.
- Ramanoël, S., Kauffmann, L., Cousin, E., Dojat, M., & Peyrin, C. (2015). Age-related differences in spatial frequency processing during scene categorization. *PloS one*, *10*(8), e0134554.
- Ramezani, F., Kheradpisheh, S. R., Thorpe, S. J., & Ghodrati, M. (2019). Object categorization in visual periphery is modulated by delayed foveal noise. *Journal*

- of vision*, 19(9), 1–1.
- Ramulu, P. (2009). Glaucoma and disability : which tasks are affected, and at what stage of disease? *Current opinion in ophthalmology*, 20(2), 92.
- 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.
- Rémy, F., Saint-Aubert, L., Bacon-Macé, N., Vayssière, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes : A life-span study. *Vision research*, 91, 36–44.
- Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision research*, 44(19), 2301–2311.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see : The need for attention to perceive changes in scenes. *Psychological science*, 8(5), 368–373.
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current opinion in neurobiology*, 12(2), 162–168.
- Rolls, E. T. (2019). The orbitofrontal cortex and emotion in health and disease, including depression. *Neuropsychologia*, 128, 14–43.
- Rosch, E., & Lloyd, B. B. (1978). Cognition and categorization.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, 8(3), 382–439.
- Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision : Effects attributable to a lossy representation in peripheral vision. *Frontiers in psychology*, 3, 13.
- Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision : It is not the stimuli but the comparison. *Journal of vision*, 10(10), 25–25.
- Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, 12(6), 852–877.
- Roux-Sibilon, A., Rutgé, F., Aptel, F., Attye, A., Guyader, N., Boucart, M., ... Peyrin, C. (2018). Scene and human face recognition in the central vision of patients with glaucoma. *PloS one*, 13(2), e0193465.
- Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019, 12). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, 19(14), 7-7.
- Rudebeck, P. H., & Rich, E. L. (2018). Orbitofrontal cortex. *Current Biology*, 28(18), R1083–R1088.
- Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images : Scaling in the woods. In *Advances in neural information processing systems* (pp. 551–558).
- Schafer, A., Rouland, J. F., Peyrin, C., Szaffarczyk, S., & Boucart, M. (2018). Glaucoma affects viewing distance for recognition of sex and facial expression. *Investigative ophthalmology & visual science*, 59(12), 4921–4928.
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features ! understanding

- recognition from the use of visual information. *Psychological science*, 13(5), 402–409.
- 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.
- Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, 26(8), 1027–1038.
- Schyns, P. G., & Oliva, A. (1999). Dr. angry and mr. smile : When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3), 243–265.
- Shulman, G. L., Sullivan, M. A., Gish, K., & Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception*, 15(3), 259–273.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst : Sustained inattentional blindness for dynamic events. *perception*, 28(9), 1059–1074.
- Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences*, 107(46), 20099–20103.
- Smith, N. D., Crabb, D., & Garway-Heath, D. (2011). An exploratory study of visual search performance in glaucoma. *Ophthalmic and Physiological Optics*, 31(3), 225–232.
- Smith, N. D., Glen, F., & Crabb, D. (2012). Eye movements during visual search in patients with glaucoma. *BMC ophthalmology*, 12(1), 45.
- Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, 14(2), 92–104.
- Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling : creating controlled scrambled images using diffeomorphic transformations. *Journal of vision*, 14(12), 6–6.
- Sun, Q., Ren, Y., Zheng, Y., Sun, M., & Zheng, Y. (2016). Superordinate level processing has priority over basic-level processing in scene gist recognition. *i-Perception*, 7(6), 2041669516681307.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262(5134), 685–688.
- Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., . . . Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, 115(35), 8835–8840.
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific reports*, 8(1), 10853.
- Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, 19(1), R11–R14.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582), 520.
- Thorpe, S., Rolls, E., & Maddison, S. (1983). The orbitofrontal cortex : neuronal activity

- in the behaving monkey. *Experimental Brain Research*, 49(1), 93–115.
- To, M. P., Carvey, K. M., Carvey, R. J., & Liu, C. H. (2019). Averaging sets of expressive faces is modulated by eccentricity. *Journal of vision*, 19(11), 2–2.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network : computation in neural systems*, 14(3), 391–412.
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339(1), 190–198.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97–136.
- Trouilloud, A., Kauffmann, L., Roux-Sibilon, A., Rossel, P., Boucart, M., Mermillod, M., & Peyrin, C. (soumis). Rapid scene categorization : From coarse peripheral vision to fine central vision.
- Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? the functional significance of the n300 component. *Neuropsychologia*, 117, 222–232.
- Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive psychology*, 15(1), 121–149.
- Ungerleider, L. G., & Haxby, J. V. (1994). ‘what’ and ‘where’ in the human brain. *Current opinion in neurobiology*, 4(2), 157–165.
- VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? is it a plane? ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30(6), 655–668.
- Vogels, R., & Orban, G. A. (1996). Coding of stimulus invariances by inferior temporal neurons. In *Progress in brain research* (Vol. 112, pp. 195–211). Elsevier.
- Wagenmakers, E.-J., Kryptos, A.-M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions : A survey of the field 33 years after loftus. *Memory & cognition*, 40(2), 145–160.
- Wallis, T. S., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., & Bethge, M. (2019). Image content is more important than bouma’s law for scene metamers. *ELife*, 8, e42512.
- Wang, P., & Cottrell, G. W. (2017). Central and peripheral vision for scene recognition : a neurocomputational modeling exploration. *Journal of vision*, 17(4), 9–9.
- Wässle, H., Grünert, U., Röhrenbeck, J., & Boycott, B. B. (1990). Retinal ganglion cell density and cortical magnification factor in the primate. *Vision research*, 30(11), 1897–1911.
- Weber, A. J., Chen, H., Hubbard, W. C., & Kaufman, P. L. (2000). Experimental glaucoma and cell size, density, and number in the primate lateral geniculate nucleus. *Investigative ophthalmology & visual science*, 41(6), 1370–1379.
- Webster, M. A. (2011). Adaptation and visual coding. *Journal of vision*, 11(5), 3–3.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, 69, 105–129.
- Williams, M. A., Baker, C. I., De Beeck, H. P. O., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature neuroscience*, 11(12), 1439.

- Woodhead, Z. V. J., Wise, R. J. S., Sereno, M., & Leech, R. (2011). Dissociation of sensitivity to spatial frequency in word and face preferential areas of the fusiform gyrus. *Cerebral Cortex*, *21*(10), 2307–2312.
- Wu, C.-T., Crouzet, S. M., Thorpe, S. J., & Fabre-Thorpe, M. (2015). At 120 msec you can spot the animal but you don't yet know it's a dog. *Journal of Cognitive Neuroscience*, *27*(1), 141–149.
- Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision : Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261.
- Yeh, S.-L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological science*, *23*(6), 608–616.
- Yucel, Y. H., Zhang, Q., Weinreb, R. N., Kaufman, P. L., & Gupta, N. (2001). Atrophy of relay neurons in magno-and parvocellular layers in the lateral geniculate nucleus in experimental glaucoma. *Investigative ophthalmology & visual science*, *42*(13), 3216–3222.