

Université de Montréal

Les attributs sous-tendant la reconnaissance d'objets visuels faits de deux composantes.

Par

Marie-Audrey Lavoie

Département de psychologie

Faculté des Arts et Sciences

Mémoire présenté en vue de l'obtention du grade de Maîtrise

en psychologie option Neurosciences Cognitives

Décembre 2021

© Marie-Audrey Lavoie, 2021

Université de Montréal

Département de psychologie, Faculté des Arts et Sciences

Ce mémoire intitulé

Les attributs sous-tendant la reconnaissance d'objets visuels faits de deux composantes.

Présenté par

Marie-Audrey Lavoie

A été évalué(e) par un jury composé des personnes suivantes

---

Président-rapporteur

Martin Arguin

Directeur de recherche

---

Membre du jury



## Résumé

La perception de la forme visuelle est le principal médiateur de la reconnaissance d'objets. S'il y a consensus sur le fait que la détection des contours et l'analyse de fréquences spatiales sont les fondements de la vision primaire, la hiérarchie visuelle et les étapes subséquentes du traitement de l'information impliquées dans la reconnaissance d'objets sont quant à elles encore méconnues. Les données empiriques disponibles et pertinentes concernant la nature des traits primitifs qu'utilise véritablement le système visuel humain sont rares et aucune ne semble être entièrement concluante. Dans le but de palier à ce manque de données empiriques, la présente étude vise la découverte des régions de l'image utilisées par des participants humains lors d'une tâche de reconnaissance d'objets.

La technique des bulles a permis de révéler les zones diagnostiques permettant de discriminer entre les huit cibles de l'étude. Les zones ayant un effet facilitateur et celles ayant un effet inhibiteur sur les performances humaines et celles d'un observateur idéal furent identifiées. Les participants n'ont pas employé la totalité de l'information disponible dans l'image, mais seulement une infime partie, ce sont principalement les segments de contours présentant une discontinuité (i.e. convexités, concavités, intersections) qui furent sélectionnés par ces derniers afin de reconnaître les cibles. L'identification des objets semble reposer sur des ensembles de caractéristiques distinctives de l'objet qui lui permettent d'être différencié des autres. Les informations les plus simples et utiles ont préséance et lorsqu'elles suffisent à mener à bien la tâche, le système visuel ne semble pas appliquer de traitement plus complexe, par exemple, l'encodage de caractéristiques plus complexes ou encore de conjonctions d'attributs simples. Cela appuie la notion voulant que le contexte influence la sélection des caractéristiques sous-tendant la reconnaissance d'objets et suggère que le type d'attributs varie en fonction de leur utilité dans un contexte donné.

Mots-clés : Reconnaissance visuelle, Perception de la forme, Psychophysique.

## **Abstract**

The main mediator of visual object recognition is shape perception. While there is a consensus that contour detection and spatial frequency analysis are the foundations of early vision, the visual hierarchy and the nature of information processing in the subsequent stages involved in object recognition, remain widely unknown. Available and relevant empirical data concerning the nature of the primitive features used by the human visual system to recognize objects are scarce and none seems to be entirely conclusive. To overcome this lack of empirical data, this study aims to determine which regions of the images are used by humans when performing an object recognition task. The Bubbles technique has revealed the diagnostic areas used by 12 adults an ideal observer, to discriminate between eight target objects. stimulus areas with a facilitatory or inhibitory effect on performance were identified.

Humans only used a small subset of the information available to recognize the targets which consisted mostly in discontinuous contour segments (i.e. convexities, concavities, intersections). Object recognition seems to rest upon contrasting sets of features which allow objects to be discriminated from one another. The simplest and most useful information seems to take precedence and it suffices to the task, the visual system does not engage in further processing involving for instance more complex features or the encoding of conjunctions of simple features. This implies that context influences the selection of features underlying human object recognition and suggests that attribute types can vary according to their utility in a given context.

Keywords : Visual recognition, Shape perception, Psychophysics

# Table des matières

Résumé .....	2
Abstract .....	3
Table des matières .....	4
Liste des tableaux .....	7
Liste des figures .....	8
Liste des sigles et abréviations .....	9
Remerciements .....	11
Introduction .....	1
Contexte théorique .....	12
L'approche cognitive .....	12
Les théories constructivistes .....	13
L'approche structurale .....	13
La théorie de reconnaissance-par-composantes .....	14
La genèse de l'approche basé sur les vues .....	16
L'approche basée sur les vues .....	17
Trois dimensions sur lesquelles se distinguent les théories structurales .....	19
La nature des traits primitifs .....	19
La stabilité des traits primitifs .....	20
La configuration des traits primitifs .....	21
La technique des Bulles .....	21
L'observateur idéal .....	22
Objectifs et Hypothèses .....	12
Méthodologie .....	25
Procédure .....	27

Tâche de reconnaissance effectuée par les humains .....	27
La tâche de l'observateur idéal .....	29
Traitement des données .....	30
La construction d'images de classification brutes. ....	30
Le calcul de l'image de classification .....	30
La transformation des images de classification individuelles brutes en score Z .....	30
Le calcul de la moyenne et de l'écart-type de référence par la technique de « bootstrapping » .....	30
Le moyennage des images de classification individuelles.....	31
Le test Pixel.....	31
Évaluation du nombre de participants partageant une région de l'image comme significativement supérieure / inférieure à 0.....	31
Corrélation des images de classification des observateurs humains vs observateur idéal :...	32
Les analyses de « zone unique » .....	32
Résultats .....	33
Performance et images de classification des participants humains .....	33
Les images de classification cumulatives .....	36
Les zones uniques .....	38
Performance et images de classification de l'observateur idéal .....	39
Observateurs humains vs observateur idéal.....	43
Discussion .....	45
Analyse comparative des IC des humains avec celles de l'observateur idéal .....	46
Caractérisation des zones diagnostiques chez l'humain.....	48
Attributs de bas niveau : contours, concavités/convexités et intersections .....	48
Attributs de niveau intermédiaire : géons, squelette et surfaces .....	51
Fragments d'image.....	52

Conclusion.....53  
Références bibliographiques .....54



## Liste des tableaux

Tableau 1 – Taux moyen de réponses correctes (%) pour les observateurs humains pour chaque objet cible lors des essais expérimentaux.....	33
Tableau 2 – Niveau de bruit moyen du masque de bruit blanc surimposé à l’image de chaque cible aux essais expérimentaux réalisés par les observateurs humains.....	33
Tableau 3 – Seuil du test Pixel pour chaque image de classification moyenne pour les observateurs humains. ....	34
Tableau 4 – Valeurs du coefficient de corrélation (r) entre les images de classification moyenne et les images de classification cumulatives pour chacune des cibles.....	38
Tableau 5 – Valeurs des coefficients de corrélation entre les zones uniques et l’IC moyenne pour chaque cible.....	39
Tableau 6 – Taux moyen de réponses correctes (%) et niveaux moyen de bruit blanc pour l’observateur idéal avec un niveau de contraste à 20 % des images des cibles. ....	40
Tableau 7 – Taux moyen de réponses correctes (%) pour l’observateur idéal avec un niveau de contraste de 50% des images des cibles.....	40
Tableau 8 – Niveaux moyens de contraste du bruit blanc pour l’observateur idéal avec un niveau de contraste de 50% des images des cibles. ....	41
Tableau 9 – Seuil du test Pixel pour chaque image de classification pour l’observateur idéal. ....	41
Tableau 10 – Valeurs du coefficient de corrélation (r) entre les images de classification moyennes des participants humains et celles de l’observateur idéal, pour chacune des cibles.....	44

## Liste des figures

Figure 1 – Les huit objets cible employés dans le cadre de cette étude et tirés de la collection de 12 objets dans l'étude de Leek, Reppa, Rodriguez & Arguin, (2009). .....	26
Figure 2 – Illustrations des stimuli employés lors de l'étude.....	27
Figure 3 – Schéma illustrant le déroulé temporel de la tâche de reconnaissance effectuée par les participants humains.....	29
Figure 4 – Images de classification moyennes pour les participants humains.....	35
.....	36
Figure 5 – Images de classifications simplifiées illustrant les zones ayant un effet facilitateur ou inhibiteur sur la performance de reconnaissance de chaque cible par les observateurs humains. .	36
.....	37
Figure 6 – Images de classification cumulatives des participants humains pour chaque cible.....	37
Figure 7 – Illustration représentant les zones uniques pour chaque cible. ....	39
Figure 8 – Images de classification moyennes pour l'observateur idéal.....	42
Figure 9 – Images de classifications simplifiées illustrant les zones ayant un effet facilitateur ou inhibiteur sur la performance de reconnaissance de chaque cible par l'observateur idéal. ....	43
Figure 10 – Images de contraste entre les IC des participants humains et de l'observateur idéal. ....	44
Figure 11 – Images de classification moyennes des cibles 1, 2, 4, 7 et 8. ....	49
Figure 12 – Image de classification moyenne de la cible 6. ....	49
Figure 13 – Images de classification moyennes des cibles 1, 2, 5 et 8 illustrant les zones facilitatrices. ....	50
Figure 14 – Image de classification moyenne de la cible 3. ....	51

## Liste des sigles et abréviations

2D: bidimensionnel

3D: tridimensionnel

dB : décibel

Hz : Hertz

IC : Image de classification

IRMf : imagerie par résonance magnétique fonctionnelles

ms : millisecondes

PNA: propriété non-accidentelle

RPC: reconnaissance par composantes

*À P-A,*

*Qui m'a appris que pour devenir un meilleur humain, il valait mieux prendre plaisir à faire les choses pour lesquelles nous étions moins doués et celles que nous pensions être les plus difficiles.*

*M-A Lavoie*

## **Remerciements**

J'aimerais remercier mon directeur Martin Arguin pour sa patience, son soutien et son encadrement ainsi que les membres de mon laboratoire qui ont su, avec leurs interactions, susciter ma curiosité et mon intérêt.

# Introduction

La perception de la forme visuelle est le principal médiateur de la reconnaissance d'objets. S'il y a consensus sur le fait que la détection des contours et l'analyse de fréquences spatiales sont les fondements de la vision primaire, la hiérarchie visuelle et les étapes subséquentes du traitement de l'information impliquées dans la reconnaissance d'objets sont quant à elles encore méconnues. L'un des objectifs fondamentaux de la recherche dans ce domaine est de déterminer les codes et les descripteurs employés par notre système visuel afin d'aboutir à une représentation cohérente des objets. La présente étude s'intéresse aux indices diagnostiques sous-tendant la perception et la reconnaissance de la forme visuelle tridimensionnelle (3D) chez l'humain. Traditionnellement, les chercheurs ont employé la méthode hypothético-déductive pour parvenir à leur fin. Des hypothèses basées sur des théories sont d'abord formulées, puis soumises à l'expérimentation. L'application rigoureuse de cette méthode n'a toutefois pas permis d'atteindre un consensus en ce qui a trait à la reconnaissance et à la représentation interne de la forme visuelle. Le but de la présente étude est de contribuer au corpus de connaissances quant aux codes et aux descripteurs par lesquels le système visuel humain représente la forme des objets tridimensionnels.

L'objectif spécifique de ce mémoire est d'effectuer un exercice visant à déterminer les parties de l'image utilisée pour identifier un objet complexe non familier fait de deux composantes. Les résultats obtenus par des participants humains lors d'une tâche de reconnaissance d'objets utilisant la technique des bulles développée par Gosselin et Schyns (2001) seront comparés à ceux obtenus par un observateur idéal accomplissant la même tâche. L'observateur idéal consiste en un programme informatique qui, dans le présent contexte, s'avère capable d'identifier, sans biais ni limitation, l'information diagnostique nécessaire à la réalisation de la tâche. Il est indispensable pour l'humain d'établir l'information contenue dans un stimulus afin de le reconnaître, il serait cependant peu efficient pour lui de prendre en charge l'entièreté de l'information disponible en raison des limites dans sa capacité d'encodage et de traitement ou encore de biais intrinsèques caractérisant les processus étudiés. Il apparaît probable que l'information employée par les humains adultes normaux pour reconnaître les stimuli diffère de celle employée par l'observateur idéal. Toutes les différences observées entre les résultats obtenus par des participants humains et ceux obtenus par l'observateur idéal régulier pourront nous éclairer sur l'information diagnostique permettant la reconnaissance d'objets 3D par des observateurs humains, normaux

# Contexte théorique

L'étude de la perception de la forme visuelle présente une longue tradition qui s'est historiquement articulée autour de grandes questions. Le traitement de la forme est-il local ou global, direct ou indirect, inné ou acquis ? Ces dichotomies portant notamment sur la nature du traitement, les conceptions de la genèse des perceptions et sur la nature des processus perceptifs ont influencé l'émergence de divers paradigmes, courants théoriques et approches qui diffèrent les uns des autres sur les plans conceptuel et contextuel.

Le premier modèle cortical du traitement de l'information visuelle fut proposé par Hubel et Wiesel (1968) et suggérait une sélectivité pour l'orientation de barres. Le traitement réalisé par les cellules ganglionnaires et celles du corps genouillé latéral et leurs champs récepteurs concentriques, situés à l'entrée du système visuel, consiste en une décomposition de l'image en minuscules unités signalant les contrastes locaux. Leur intégration donne ensuite naissance aux champs récepteurs des cellules de l'aire V1, qui sont dotées d'une sélectivité pour l'orientation et la fréquence spatiale. Cependant, ces unités de traitement représentent elles aussi l'information visuelle de manière extrêmement locale. Il semble donc nécessaire de supposer un traitement subséquent destiné à construire des représentations plus globales de la forme des objets.

## L'approche cognitive

La science cognitive a vu le jour lors d'un symposium portant sur la théorie de l'information au MIT en 1956. Cet événement avait la particularité de rassembler des chercheurs provenant de disciplines variées qui employaient un langage provenant des systèmes de communication afin de décrire la perception et l'activité mentale. Fortement influencés par l'avènement des sciences informatiques, les tenants de cette approche avancent que le traitement de l'information sensorielle se fait par analyse et décomposition des processus mentaux en appliquant des ensembles de règles précises. L'explication des données comportementales dans cette perspective requiert des structures internes permettant d'encoder et de transformer l'information captée dans l'environnement. L'activité cognitive s'appuie sur des représentations mentales, décrites à l'aide de schémas et de règles. Le contenu des représentations y est envisagé comme le produit d'une suite de calculs internes ressemblant à ceux réalisés par un ordinateur.

Une théorie computationnelle doit être en mesure d'identifier l'information visuelle qui doit être représentée et d'expliquer en quoi elle est utile. L'élaborer exige donc, des informations sur la manière par laquelle le cerveau construit des représentations du monde et des objets à partir de l'image rétinienne et des règles de correspondance entre cette image, qui est bidimensionnelle, et la structure tridimensionnelle de l'environnement.

## **Les théories constructivistes**

Selon les modèles constructivistes, la perception est un processus actif impliquant la construction d'une représentation interne de l'environnement à l'aide du traitement de l'information captée par les récepteurs sensoriels. Dans cette optique, la perception est analogue à la formulation d'hypothèses sur la composition et l'organisation du monde et à ce titre, sa véracité n'est pas garantie. Dans le contexte particulier de la reconnaissance d'objets, des représentations internes des objets familiers doivent être maintenues en mémoire et la mise en relation du stimulus perçu avec la représentation mnésique correspondante est requise pour qu'il puisse être reconnu correctement.

Jusqu'à tout récemment, deux grands courants de pensée ont dominé le domaine d'étude de la représentation de la forme visuelle 3D. Il s'agit de l'approche structurale et de l'approche basée sur les vues.

## **L'approche structurale**

Le modèle de la perception d'objets visuels 3D proposé par Marr et Nishihara (1978) postule que la représentation interne d'un objet culmine en un code reflétant les parties constituantes de l'objet sous forme de cylindres généralisés et les relations entre ces parties. S'appuyant sur ce modèle, Biederman (1987) mit au point la théorie de la reconnaissance par composantes (RPC), qui demeure l'instanciation des théories structurales la plus largement acceptée. Les alternatives se distinguant principalement, mais non exclusivement, sur la nature des traits primitifs proposés, il paraît donc approprié, d'un point de vue historique, d'utiliser la théorie RPC à titre d'exemple pour illustrer l'approche structurale.



## La théorie de reconnaissance-par-composantes

La RPC postule que la représentation visuelle des objets solides et rigides dans notre environnement consiste en un assemblage d'éléments volumétriques simples nommés géons, une contraction de l'expression « geometric ions » (ions géométriques). Ces unités élémentaires constituent l'alphabet visuel servant à la représentation des parties constituantes des objets 3D. Les géons auxquels les parties d'objets complexes sont assimilés, sont eux-mêmes définis selon une collection de propriétés de l'image 2D des contours internes et externes de l'objet qui sont dites non accidentelles. Ces propriétés non-accidentelles (PNA) sont la cotermination, la symétrie, la courbure, le parallélisme et la colinéarité (Biederman, 1987). Leur valeur « non-accidentelle » résulte de leur capacité à signaler de manière invariante et fiable (excluant les points de vue accidentels<sup>1</sup>) certaines propriétés de la forme tridimensionnelle de l'objet. La représentation des objets est donc constituée de géons, qui eux-mêmes sont définis par une conjonction particulière de PNA, rendant ainsi la représentation invariante aux rotations en profondeur. Enfin, la détermination des relations spatiales entre les géons surviendrait une fois leur encodage terminé. Ce traitement donnerait donc naissance à une représentation des objets tridimensionnels invariante au point de vue, c'est-à-dire qui est stable peu importe le point de vue. L'importance de cette dernière caractéristique tient au fait qu'elle permet d'assurer la constance de forme, soit la capacité de reconnaître deux formes 3D comme identiques malgré des différences majeures dans les projections rétinienne qui résultent de points de vue différents. Les portions des contours permettant de générer les géons ont en commun l'invariance au point de vue et peuvent être déterminées avec quelques points sur chaque segment. Cela permettrait l'extraction d'un trait primitif avec une grande résistance aux variations du point de vue, au bruit et à l'occlusion. (Biederman, 1987)

Les fondements de la RPC reposent donc sur trois concepts, soit la segmentation de l'objet en ses parties, le caractère hiérarchique du traitement effectué et l'indépendance de la configuration spatiale des géons.

---

<sup>1</sup> Points de vue accidentels : point de vue générant des régularités au niveau de l'image visuelle ayant une faible probabilité de survenir dans la nature (Biederman, 1987)

### La segmentation de l'objet en ses parties

Les résultats obtenus lors d'une première série d'expériences rapportées par Biederman et Ju, (1988) indiquent que des objets représentés par un dessin étaient reconnus aussi bien et rapidement que lors de la présentation de photographies haute définition des mêmes objets. Cela a permis d'utiliser des dessins d'objets comme stimuli dans l'étude de la perception et la reconnaissance de la forme visuelle.

Dans une autre expérience, les participants devaient reconnaître des dessins d'objets dont 50% du contour était manquant. Les contours manquants étaient situés soit au milieu de segments droits ou avec une courbure stable ou encore à l'intersection de segments. La performance des participants était nettement meilleure dans la première que dans la seconde condition. Selon la logique de la RPC, l'absence de jonctions de contours, qui sont des indices structuraux importants, empêche l'extraction des géons, rendant ainsi la reconnaissance de l'objet laborieuse et dans certains cas impossible (Biederman, 1987). L'étude d'Arguin et Saumier, (2004), interprète le coût qu'entraîne le partage des parties d'un objet cible avec des distracteurs lors d'une tâche de recherche visuelle comme un appui à l'hypothèse selon laquelle les formes complexes sont représentées en termes des parties qui les constituent.

### Le Traitement hiérarchique

Biederman, (1987) propose que le traitement sous-tendant la reconnaissance d'objets est organisé de manière hiérarchique. D'abord survient l'extraction des contours à partir des contrastes de luminance, texture et/ou couleur. Ensuite deux processus distincts surviendraient simultanément ; la segmentation de l'objet en ses parties et la détection des propriétés non-accidentelles de leurs contours. C'est à travers ces processus séquentiels qu'il devient possible d'assimiler les parties de l'objet à des géons, dont l'arrangement pourra ensuite être établi. C'est cette représentation, faite d'une collection de géons et de leur assemblage, qui pourra être appariée à la représentation de l'objet en mémoire en vue de le reconnaître.

### L'indépendance de la configuration spatiale

Ce terme signifie que la représentation de la configuration des composantes d'un objet est indépendante de la représentation des composantes elles-mêmes. Les résultats obtenus dans l'étude d'Arguin et Saumier (2004) indiquent que le partage des composantes d'un objet et celui de leur

organisation spatiale affectent tous deux les performances des participants lors d'une tâche de recherche visuelle. Ils démontrent également que les effets délétères observés de chacun de ces facteurs sont purement additifs. Le partage des parties et celui de l'organisation spatiale des cibles avec celles des distracteurs entraînent chacun une augmentation des temps de réponses lors de la tâche de recherche visuelle ainsi qu'une augmentation des pentes des temps de réponse en fonction du nombre d'éléments affichés. Dans le contexte d'une tâche de recherche visuelle accélérée comme celle effectuée dans cette étude, le coût qu'entraîne le partage des parties entre les cibles et les distracteurs appuie l'hypothèse selon laquelle les formes complexes seraient analysées sur la base des parties qui les constituent. Le fait que le partage de l'organisation spatiale de la cible avec celle des distracteurs entraîne un coût élevé sur les performances, et ce même lorsque les objets sont faits de parties différentes, démontre l'importance de l'organisation spatiale des parties constitutives dans la représentation des objets complexes. Le fait que les effets observés qu'engendre l'un et l'autre de ces facteurs sont additifs indique que les deux mécanismes en jeu sont indépendants et donc qu'ils ne s'influencent pas l'un l'autre. Cela entre en conflit avec l'idée que la représentation de formes visuelles repose seulement sur des informations globales de l'objet ainsi qu'avec celle voulant qu'elle repose uniquement sur l'information locale provenant des contours de l'objet. (Cortese & Dyre, 1996; Edelman & Weinshall, 1991; Lades & al., 1993; Poggio & Edelman, 1990; Ullman, 1989).

### **La genèse de l'approche basée sur les vues**

L'étude de Biederman et Gerhardstein (1993) emploie un paradigme d'amorçage et comporte au total cinq expériences. À l'aide de diverses tâches d'identification, ils s'intéressent à l'effet de la rotation en profondeur (i.e. rotation de l'objet autour d'un axe vertical; couramment qualifiée de changement de point de vue) sur la reconnaissance. Les expériences 1 et 2 ont été conçues dans le but d'examiner l'invariance au point de vue à l'aide de stimuli ayant une faible propension à présenter des parties différentes d'une vue à l'autre. Les résultats obtenus indiquent une robuste invariance à la rotation en profondeur pour ce qui est des objets familiers présentés. Cette invariance est toutefois dépendante des trois conditions suivantes : les objets doivent posséder une description structurale géonique (condition 1), qui doit être distincte entre les objets présentés (condition 2) et qui doit être maintenue sous différents points de vue (condition 3). L'expérience 3 à la différence des deux premières, emploie des objets non familiers lors d'une tâche

d'appariement (pareil/différent). Une proportion des essais remplit les trois conditions nécessaires à l'invariance au point de vue alors que les autres remplissent les deux premières conditions, mais contreviennent à la troisième en raison d'une occlusion ou de l'accrétion d'un géon. Les stimuli du premier groupe n'affichaient pratiquement aucun effet de rotation en profondeur. Donc lorsqu'ils sont considérés individuellement, les géons semblent spontanément perçus comme invariants au point de vue. L'expérience 4 démontre que présenté seul, un géon est effectivement invariant à la disparité d'orientation. L'expérience 5 démontre que l'ajout d'un géon significatif est suffisant pour conférer l'invariance au point de vue aux stimuli d'Eldelman et al. (1989); objets en forme de « trombones tordus », dont la reconnaissance est autrement hautement dépendante au point de vue (Biederman & Gerhardstein, 1993).

Dans la théorie RPC (Biederman, 1987), le concept des traits primitifs volumétriques a comme fonction d'expliquer la constance de forme qui, dans l'esprit de Biederman, semblait être synonyme de performance invariante au point de vue. Or, tel qu'expliqué dans la section suivante, les deux concepts doivent être distingués et plusieurs études ont démontré que malgré la constance de forme, l'invariance au point de vue s'est avérée impossible à démontrer empiriquement. Ce dernier fait est celui ayant mené à l'élaboration de l'approche basée sur les vues.

## **L'approche basée sur les vues**

Alors que la théorie RPC implique des représentations d'objets qui sont entièrement invariantes au point de vue (hormis ceux dits 'accidentels'), de nombreuses études ont révélé que le système visuel humain présente des performances de reconnaissance d'objet qui sont affectées par le changement de point de vue (e.g. Tarr & Bülthoff, 1995; Hayward & Tarr, 1997; Tarr, Williams, Hayward & Gauthier, 1998; Hayward & Williams, 2000). Cette règle s'applique même dans le cas d'objets élémentaires assimilables à des géons. Ainsi, une étude s'étant avérée particulièrement décisive dans le débat concernant l'invariance au point de vue de la performance perceptive chez l'humain est celle de Tarr et al. (1998). Ils élaborèrent une série d'expériences afin de mettre à l'épreuve le postulat de la RPC selon lequel les géons seraient invariants au point de vue (Biederman, 1987). La première d'entre elle emploie une tâche d'appariement (pareil/différent) et deux versions d'images représentant des géons (dessinés au trait ou images ombragées). Deux images étaient présentées consécutivement sous des points de vue différant de 0°, 45°, 90°. Le participant devait déterminer s'il s'agissait du même géon ou s'il s'agissait d'objets différents. Les

auteurs rapportent d'importants effets de rotation, même pour des objets constitués d'un unique géon et qui respectaient les conditions pour l'invariance au point de vue posées précédemment par Biederman et Gerhardstein (1993). De telles observations ont amené certains auteurs à rejeter la RPC et à proposer une théorie alternative appelée l'approche basée sur les vues.

Contrairement aux modèles de l'approche structurale, ceux basés sur les vues ne formulent aucune hypothèse explicite quant aux caractéristiques sous-tendant la représentation de la forme visuelle, hormis l'affirmation que ces caractéristiques sont strictement 2D (i.e. elles n'encodent la profondeur d'aucune manière; (e.g., Poggio & Edelman, 1990; Ullman, 1989; Basri & Ullman, 1991). Partant de ce postulat fondamental, la constance de forme est expliquée par des processus tels que l'interpolation, la rotation mentale, l'alignement des descriptions picturales (i.e. 2D), etc. Spécifiquement, c'est la mise en branle de ces opérations qui permettrait l'appariement entre l'image d'un objet connu observé sous un point de vue non-familier et des images de ce même objet sous des points de vue familiers maintenues en mémoire (Hayward & Tarr, 1997; Tarr, Williams, Hayward & Gauthier, 1998; Edelman & Bulthoff, 1992; Poggio & Edelman, 1990; Basri & Ullman, 1991; Shepard & Metzler, 1971; Tarr & Pinker, 1989; Ullman, 1989).

À l'appui de l'approche basée sur les vues, certaines études employant des modèles d'inspiration biologique de réseaux neuronaux hiérarchiques démontrent que le traitement de l'image 2D à lui seul suffit afin d'identifier un objet de façon rapide et précise (Vetter, Hurlbert & Poggio, 1995). L'approche des vues arrive ainsi à éviter les calculs complexes requis pour l'inférence des structures 3D. En résumé, l'approche des vues est appuyée par des expériences psychophysiques démontrant que l'indépendance au point de vue peut être acquise à travers l'expérience de points de vue variés (Tarr & Pinker 1989, Bulthoff, Edelman & Tarr, 1995; Hayward & Tarr, 1997; Tarr, Williams, Hayward & Gauthier, 1998) et sur d'autres études démontrant que le traitement de l'image 2D permet une identification précise et rapide des objets (Riesenhuber & Poggio, 1999; Serre, Oliva & Poggio, 2007).

Tel qu'indiqué précédemment, il existe une variété de théories structurales de la perception de la forme, se distinguant les unes des autres sur trois grande dimensions, soit; la nature des traits primitifs, leurs stabilité et leur configuration, qu'il importe de passer en revue.

## Trois dimensions sur lesquelles se distinguent les théories structurales

### La nature des traits primitifs

L'ensemble des théories considère les segments de contours, leurs intersections et le contraste comme des caractéristiques pertinentes pour la reconnaissance d'objets. Celles-ci sont cependant considérées comme un moyen par lequel des représentations plus complexes et utiles sont construites plutôt que comme une finalité. Selon plusieurs, les concavités marquées, notamment ce qu'on appelle les minima négatifs de courbure, sont priorisées étant donné leur importance pour la segmentation de l'objet en ses parties constituantes. (Breakspears, Brammer, Das & Williams, 2004; Smith, Singh, Williams & Greenlee, 2001 ; Uyar, Shomstein, Greenberg & Behrmann 2016; Breakspear, Brammer & Robinson, 2003). Par ailleurs, les contours convexes seraient utiles pour caractériser la forme de ces parties (Bertramini & Wagemans, 2013).

Bien qu'elles postulent toutes que la perception de la forme visuelle repose sur une segmentation de l'objet en plus petites unités (Marr & Nishihara, 1978; Marr, 1982; Biederman, 1987; Pentland, 1986; Feldman & al., 2013; Leyton, 1989; Leek, Reppa & Arguin, 2005; Ullman, Vidal-Naquet & Sali, 2002), ces théories se distinguent quant à la nature exacte de ces unités de représentation.

Marr (1982) avance que les contours servent à la construction d'une représentation globale en relief de 2 1/2D (« full primal sketch »). Les parties de l'objet seraient quant à elles encodées en tant que cylindres généralisés. Biederman (1987) propose quant à lui que l'extraction des propriétés 2D non accidentelles (PNA) des contours mène à la représentation de ses parties en termes de géons. La représentation d'objets complexes nécessiterait que les relations spatiales entre les géons soient aussi représentées. Pentland (1986) propose une théorie semblable à celle de Biederman au sens où les représentations de la forme reposeraient sur des traits primitifs volumétriques. Toutefois, au lieu de géons il fait appel à une famille de formes paramétrées connues sous le nom de superquadratiques. Feldman et al. (2013) et Leyton (1989) proposent quant à eux que les descriptions de la forme reposent sur l'agencement de parties allongées. Les axes médians de ces parties et leurs relations spatiales forment le squelette de l'objet. Le contour serait la source d'information dans l'image qui permettrait d'inférer les axes médians. Dans cette optique le squelette sous-jacent de la forme serait le produit de l'agencement des axes médians des parties

de l'objet telles que segmentées aux minima négatifs de courbure. Leek, Reppa et Arguin (2005) avancent que la représentation des objets reposerait sur la collection de leurs surfaces visibles et de leurs relations spatiales. Ces derniers ont examiné l'hypothèse selon laquelle la forme des objets est représentée par des composantes volumétriques et leurs configurations dans trois expériences employant un paradigme d'appariement tout-à-partie (« whole-part matching ») dans laquelle les participants devaient appairer des objets entiers non-familiers représentés en termes de leurs contours à des images faites d'un sous-ensemble de ces contours. Les résultats des expériences 1 et 2 montrent de meilleures performances lorsque les parties d'objets constituent des composantes volumétriques (tels des géons) ou des surfaces fermées que si elles sont faites de collections de contours interconnectés qui ne définissent ni un volume ni une surface fermée. Aucune différence n'est observée entre les conditions de composantes volumétriques ou de surfaces fermées. Ainsi la notion de « surface visible de la forme » pourrait être envisagée comme trait primitif. Enfin, Ullman, Vidal-Naquet et Sali (2002) avancent que les attributs fondamentaux sont les fragments (portions d'image de complexité intermédiaire présentant une valeur informative élevée pour la classe d'objets à reconnaître) et les « caractéristiques étendues », qui sont constituées d'images de ce qu'ils appellent les composantes (i.e. les parties de l'objet sélectionnés comme fragments) captées sous des points de vue variés.

### **La stabilité des traits primitifs**

Les auteurs qui spécifient la nature des traits primitifs de l'objet sont plutôt enclins à supposer que ces derniers sont stables. Ce raisonnement découle d'un postulat implicite selon lequel, l'information contenue dans la représentation d'objets qui est utilisée lors de la reconnaissance est exhaustive. Certaines de ces théories supposent qu'il existe une classe particulière de traits primitifs qui permettrait de se représenter les objets de façon stable, sans égard au point de vue ou au contexte dans lequel la tâche est réalisée (Biederman, 1982; Andrews, Clarke, Pell & Hartley, 2010; Schyns & Oliva, 1994; Sekuler & Bennett, 2001; Smith, Singh, Williams & al., 2001).

Cependant, ce postulat n'est pas obligatoire. Il serait par exemple possible d'envisager les traits primitifs proposés par une théorie comme une collection d'éléments disponibles et dont l'usage varierait en fonction du contexte. Plus précisément, cette utilisation serait fonction de leur utilité dans un contexte donné et celle-ci ne serait pas tenue de demeurer fixe à travers les points de vue (Coggan, Baker & Andrews, 2019; Daugman, 1989; Wiesmann, Caplette, Willenbockel, Gosselin

& Võ, 2021; Willenbockel, Sadr & Fiset, 2010; Ullman, 2007; Ullman, Assif, Fetaya & Harari, 2016; Ullman & Bart, 2004; Harel, Ullman, Epshtein & Bentin, 2007; Harel, Ullman, Harari & Bentin, 2011). Ainsi, les caractéristiques employées en certaines circonstances en raison de leur contribution à la reconnaissance pourraient être ignorées dans des contextes différents où elles n'auraient aucune valeur diagnostique.

La proposition la plus évidente en ce qui concerne la variabilité des traits fut faite par les tenants de l'approche du traitement de l'information visuelle basée sur les catégories informatives de fragments. Les fragments seraient sélectionnés en fonction de leur valeur informative afin d'accomplir une tâche donnée (Ullman, 2007; Ullman & Bart, 2004; Ullman, Assif, Fetaya & Harari, 2016; Harel, Ullman, Harari & Bentin, 2011; Harel, Ullman, Epshtein & Bentin, 2007).

### **La configuration des traits primitifs**

Certains chercheurs avancent que les conjonctions de propriétés sont utilisées afin d'encoder les relations spatiales entre les traits primitifs caractérisant un objet, donc en faveur de l'argument de l'encodage structural (Barenholtz & Tarr, 2011; Biederman, 1987). L'étude d'Arguin et Saumier (2004), décrite précédemment, démontre l'existence d'un mécanisme dédié à la représentation de la configuration des parties constituant les objets complexes. Cette observation constitue un argument de poids à l'appui d'une implication obligatoire du traitement configural dans la reconnaissance d'objets.

Néanmoins, d'autres auteurs avancent plutôt que les relations spatiales à l'intérieur de l'objet ne sont pas pertinentes et n'ont par conséquent aucun besoin d'être encodées explicitement (Crouzet & Serre, 2011; Serre, Oliva & Poggio, 2007).

### **La technique des Bulles**

La technique des bulles qui fut développée par Gosselin et Schyns (2001) est une technique d'images de classification, dans laquelle un masque opaque percé d'ouvertures gaussiennes<sup>2</sup> et dont la localisation est aléatoire, révèle certaines parties de la cible située derrière le masque. L'aspect

---

<sup>2</sup> À l'intérieur de chaque ouverture composant le masque, la révélation du contraste de l'image de la cible est maximale au centre. Ce taux diminue graduellement vers la périphérie de l'ouverture en suivant une fonction gaussienne.



gaussien des ouvertures a pour effet d'éliminer les artéfacts de bordure qui pourraient autrement être produits par le masque faisant occlusion. La technique d'analyse utilisée afin de construire les images de classification repose sur le calcul de la corrélation entre la visibilité de chaque point de la cible et la performance des participants. Cette corrélation indique l'efficacité avec laquelle chacune des régions de l'image cible contribue à la réalisation correcte de la tâche. Cette technique a servi dans plusieurs contextes, notamment pour l'étude de la perception des visages, à laquelle elle a apporté une contribution importante (Blais & al., 2012; Gosselin & Schyns, 2001; Willenbockel & al., 2010). La technique des bulles fut également utilisée dans bon nombre d'études comportementales portant sur la catégorisation de scènes naturelles (McCotter, Gosselin, Sowden, & Schyns, 2005) et sur le changement perceptuel d'images ambiguës (Schyns, Bonnar, & Gosselin, 2002).

Précédemment il fut question d'une panoplie d'approches théoriques cherchant à expliquer les mécanismes impliqués dans la reconnaissance d'objets. L'un des aspects fondamentaux qui distingue les diverses approches théoriques concerne la nature des propriétés de l'image qui sous-tendent la représentation que l'humain se crée de la forme des objets. Ultimement, les autres dimensions distinguant les théories reposent en grande partie sur le choix des traits primitifs utilisés pour constituer la représentation.

Présentement, les données empiriques disponibles et pertinentes concernant la nature des traits primitifs qu'utilise véritablement le système visuel humain sont rares et aucune ne semble être entièrement concluante. Dans l'objectif de pallier ce manque de données empiriques, la présente étude vise la découverte des régions de l'image utilisées par des participants humains lors d'une tâche de reconnaissance d'objets. En révélant les régions de l'image qui sont diagnostiques pour la reconnaissance d'objets par des observateurs humains, nous espérons être en mesure de faire des inférences pertinentes quant aux types de traits primitifs sous-tendant leur représentation de la forme visuelle. Nous utiliserons à cette fin la technique des bulles dans une tâche de reconnaissance d'objets réalisée par des participants humains normaux.

## **L'observateur idéal**

Un observateur idéal est un dispositif informatique capable d'effectuer une tâche perceptive quelconque de manière optimale en traitant la totalité des informations disponibles, soit sans contraintes ou avec des contraintes spécifiées (Geisler, 2011). L'observateur idéal possède

plusieurs utilités dans le champ de la recherche en vision, telles que l'identification des propriétés des stimuli pertinents à la réalisation d'une tâche (p. ex. déterminer quelles parties d'un objet permet d'identifier celui-ci) et la description de la manière dont ces propriétés sont utilisées pour la réalisation de la tâche. Il peut également servir de comparatif lorsqu'utilisé de concert avec des participants humains ou animaux réalisant la même tâche perceptuelle. La mise en commun des informations obtenues grâce à l'utilisation de l'observateur idéal permet également de produire des hypothèses quant au fonctionnement du système perceptuel (ici, la vision) et peut servir de modèle du fonctionnement sensoriel (Geisler, 2011). Dans le contexte de ce mémoire, l'observateur idéal fournira donc un point de référence auquel il sera possible de comparer les performances de systèmes visuels réels comme celui de participants humains.

L'observateur idéal est basé sur les théories probabilistes, les statistiques et la théorie de l'information et peut prendre plusieurs formes. Dans le cas présent, l'observateur idéal recevra les mêmes informations que les participants humains, c'est-à-dire des objets cachés par un masque percé d'ouvertures gaussiennes, telles que décrites dans la section précédente. Puis, il effectue des corrélations entre l'information présente à l'intérieur des bulles et les huit objets cible. La corrélation la plus élevée dicte la réponse qui sera fournie par l'algorithme. De la même manière qu'avec les participants humains, des images de classification seront calculées afin de déterminer les régions des stimuli qui facilitent ou inhibent l'identification des cibles. Ces résultats seront ensuite comparés à ceux des observateurs humains. Les différences seront attribuables aux limitations et biais perceptifs du système visuel humain. Ces différences contribueront à notre capacité à inférer la nature des traits primitifs utilisés par ce dernier pour la représentation des objets employés dans l'expérience.

## Objectifs et Hypothèses

L'objectif de ce mémoire est de déterminer quelles régions de l'image sont utiles aux observateurs humains pour la reconnaissance d'objets 3D non familiers faits de deux composantes. Il s'agira ainsi d'identifier les zones diagnostiques permettant de discriminer entre huit images cibles. Plus spécifiquement, nous voulons identifier les zones ayant un effet facilitateur et celles ayant un effet inhibiteur sur les performances humaines et celles d'un observateur idéal lors d'une tâche de reconnaissance d'objets visuels.

Il est attendu que la technique des bulles révèle certaines régions de chaque image cible comme ayant été plus ou moins utiles que d'autres. C'est donc dire que les diverses parties d'une image ne contribueront pas toutes également à la reconnaissance de l'objet qui y est illustré. À cet égard, il semble y avoir consensus à l'effet que la discontinuité des contours (ex. convexités, concavités, intersections des contours) possède une valeur informative particulièrement élevée afin de spécifier la forme des objets (e.g. Biederman, 1987). Ceci mène à la prédiction que les contours internes ou externes des objets affichant une quelconque discontinuité seront privilégiés par rapport aux contours droits ne présentant pas d'intersections.

Enfin on s'attend à ce que les observateurs humains utilisent les régions des images qui permettent effectivement de discriminer entre les objets cibles dans le contexte d'une tâche de reconnaissance. Cependant il est également attendu qu'ils ne seront pas en mesure d'utiliser la totalité de l'information discriminante disponible en raison des limites intrinsèques du système visuel humain (e.g. limite d'acuité, courbe de la sensibilité au contraste, etc.) et de biais possibles. Cela nous amène à prédire que les régions de l'image utilisées efficacement par les humains pour la reconnaissance d'objets devraient constituer un sous-ensemble de l'information discriminante révélé par l'observateur idéal.

## Méthodologie

**Participants :** Un groupe de 12 participants adultes, sains, dont la vision est normale ou corrigée (20/20) et âgés de 18 à 35 ans ont été recrutés via le site web du Laboratoire de neurocognition visuelle de l'Université de Montréal (<https://www.neurocognitionvisionlab.org>). Tous ont fourni leur consentement libre et éclairé et le projet a été approuvé par le comité d'éthique de la recherche en éducation et psychologie de l'Université de Montréal. Le plan de mitigation des risques de transmission de la COVID19 tel qu'approuvé par le Comité de reprise des activités de recherche du Bureau de la recherche, développement et valorisation de l'Université de Montréal a été suivi intégralement lors des rencontres avec les participants.

**Matériel :** L'expérience fut développée avec le logiciel MATLAB (©1994-2017 The MathWorks, Inc.) et la Psychophysics toolbox (Kleiner, Brainard, Pello, Ingling, Murray & Broussard, 2007), permettant de construire des tests psychophysiques. Les données ont été compilées dans des fichiers MATLAB. L'expérience fut effectuée sur un ordinateur HPZ230 avec un système d'exploitation Windows 10 et équipé d'une carte graphique NVIDIA GeForce GTX970. La présentation des stimuli fut réalisée sur un écran HD ASUS VG248QR de 24 pouces et un taux de rafraîchissement de 60Hz situé à une distance de 57 cm du visage du participant. La position de la tête de l'observateur était maintenue en place avec l'aide d'une mentonnière.

**Stimuli :** Les images des objets servant de cibles furent générées à l'aide du logiciel 3DSMax (© 2017 Autodesk Inc.) avec un rendu en perspective faisant usage d'une seule source lumineuse. Tous les objets peuvent être décomposés en deux composantes selon les principes de la RPC, ont une surface uniforme et achromatique et sont présentés sur un fond blanc. L'orientation 3D selon laquelle les cibles ont été présentés dans l'expérience correspond à celle de la Fig. 1. Ainsi, l'axe principal de tous les objets est approximativement aligné dans la même direction. Les hauteurs et largeurs minimales et maximales des images des cibles sont les suivantes : hauteur, de 9,5 à 10 degrés d'angle visuel; largeur, de 6,5 à 10 degrés d'angle visuel. Ces objets constituent un sous-ensemble de ceux ayant servi à l'étude de Leek, Reppa, Rodriguez et Arguin (2009) recouverts d'une texture achromatique et sont fait de deux composantes. Des concavités facilement visibles permettent d'identifier clairement la jonction des deux composantes constituant chaque objet. Un chiffre de 1 à 8 assigné à chaque cible fut utilisé pour les identifier.

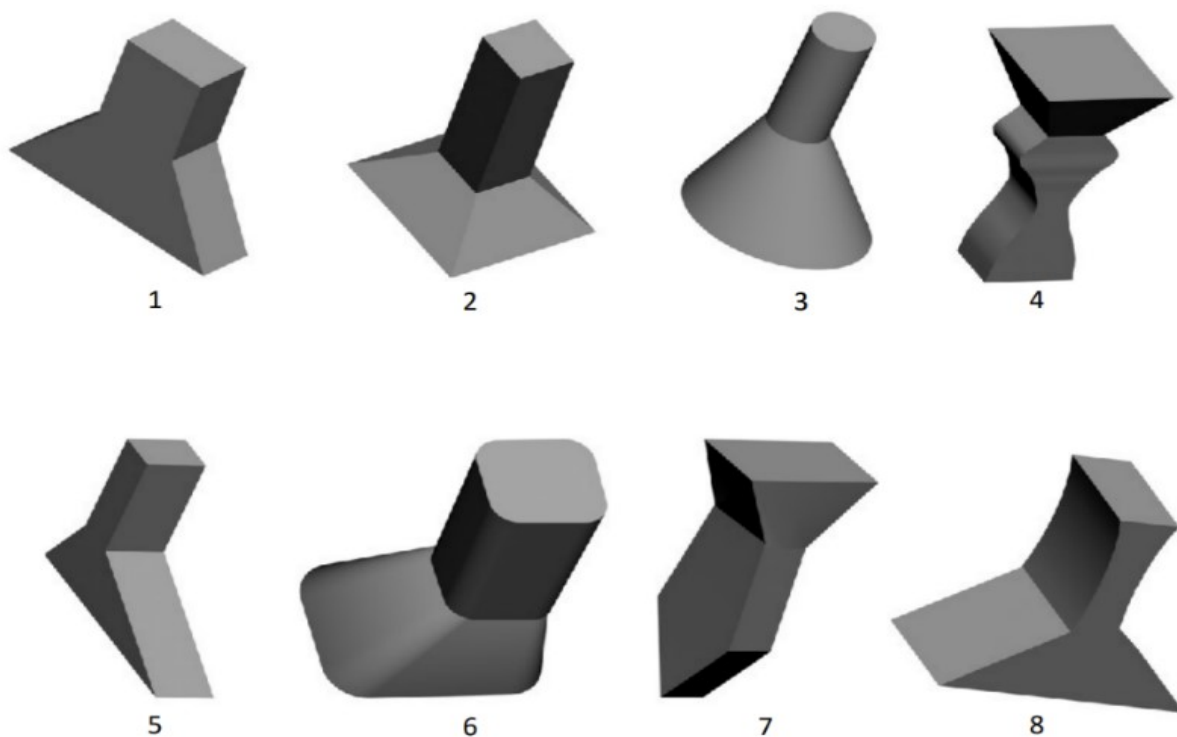
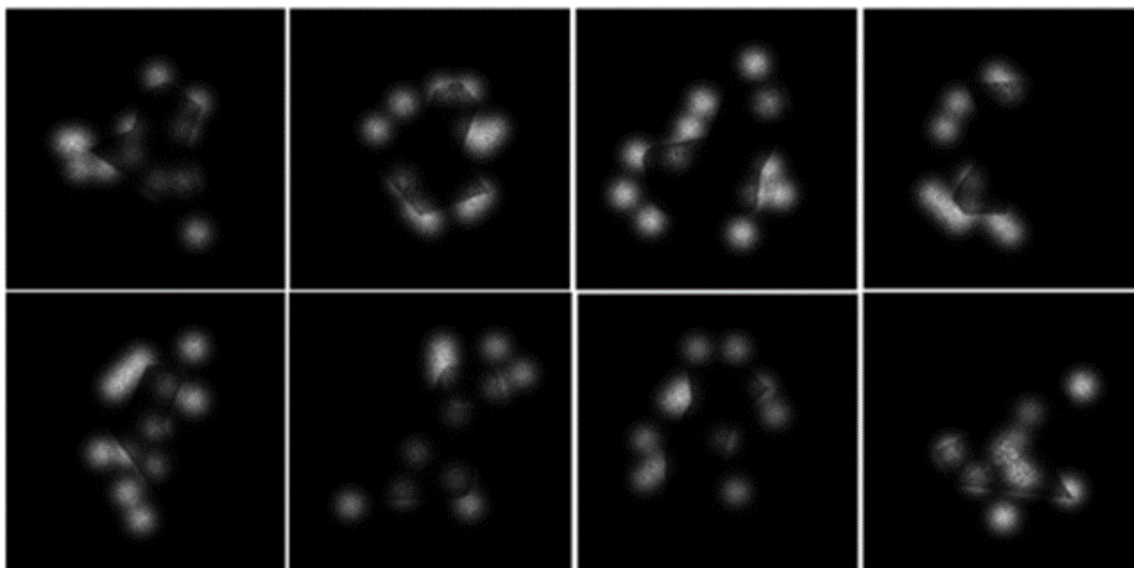


Figure 1 – Les huit objets cible employés dans le cadre de cette étude et tirés de la collection de 12 objets dans l'étude de Leek, Reppa, Rodriguez & Arguin, (2009).

À chaque essai, un masque de bruit blanc était superposé à la cible et le niveau de contraste de ce masque était ajusté afin de maintenir la performance des participants aux environs de 50% de réponses correctes. Un nouveau masque était généré à chaque essai et les règles déterminant son niveau de contraste sont décrites plus bas.

Finalement, un masque d'occlusion était superposé à l'image de la cible + bruit blanc. Le masque était noir et percé de 12 ouvertures gaussiennes circulaires d'une taille de 0,93 degrés d'angle visuel. La localisation de ces ouvertures était déterminée aléatoirement à chaque essai. La Fig. 2 donne des exemples de stimuli tels que présentés aux participants lors de la phase expérimentale.



*Fig 2. Exemples de stimuli composés de l'image d'une cible sur laquelle est superposé un masque de bruit blanc et le masque d'occlusion percé d'ouvertures révélant partiellement la cible. Un exemple est présenté pour chacune des cibles utilisées.*

Figure 2 – Illustrations des stimuli employés lors de l'étude

## **Procédure**

### **Tâche de reconnaissance effectuée par les humains**

Les observateurs humains devaient effectuer une tâche de reconnaissance d'objets qui étaient partiellement révélés, tel qu'illustré à la Fig. 2. Ils avaient reçu la consigne de nommer le chiffre assigné à la cible lors d'une phase de familiarisation et une phase d'apprentissage. L'expérimentateur entrait alors les réponses fournies par les participants en appuyant sur la touche d'un clavier filaire correspondant au numéro donné par le participant à chaque essai.

#### La phase de familiarisation

Dans un premier temps, l'ensemble des participants devait se familiariser avec les stimuli de l'étude. Une feuille contenant les huit stimuli et les numéros leur étant associés était mise à leur disposition à cette fin. Les participants avaient reçu la consigne d'informer l'expérimentateur lorsqu'ils jugeaient être prêts à commencer l'expérience, après quoi la phase pratique débutait immédiatement.

### La phase de pratique

Cette phase avait lieu lors de la première séance ou lorsque plus d'une semaine s'était écoulée entre deux séances pour un participant donné. Les huit stimuli y étaient présentés individuellement en ordre aléatoire avec un contraste élevé, sans avoir subi de dégradation et demeuraient visibles jusqu'à l'enregistrement de la réponse du participant par l'expérimentateur. Un signal sonore et visuel l'informait de la validité des réponses fournies. Ainsi, le chiffre associé à l'objet apparaissait à l'écran et un son grave lui indiquait s'il avait fourni une réponse incorrecte alors qu'un son aigu signifiait qu'il avait fourni une réponse correcte. Chaque objet devait être associé avec succès au chiffre lui correspondant, à deux reprises consécutives, afin que la phase pratique prenne automatiquement fin.

### La phase expérimentale

Cette phase commençait immédiatement après la phase de pratique. Elle était constituée de huit séances de tests d'une durée d'environ une heure chacune. Chaque séance était composée de deux blocs expérimentaux comprenant 280 essais chacun durant lesquels chaque cible y était présentée 35 fois, pour nombre total de 4480 essais par participant. La tâche consistait à identifier les objets présentés. Chaque essai débutait avec l'apparition d'une croix de fixation au centre de l'écran. La cible était ensuite présentée au même endroit. Le contraste de l'image des cibles était réduit à 20% de sa valeur initiale (telle qu'illustrée à la Fig. 1). Cette réduction de contraste s'est avérée nécessaire car autrement, la performance des participants était extrêmement élevée, même lorsque le contraste du masque de bruit blanc qui y était superposé (voir paragraphe suivant) était maximal. C'est sur la base d'une étude pilote que la réduction du contraste de l'image des cibles a été fixée à 20%.

Les stimuli étaient masqués par un bruit blanc dont le niveau de contraste était ajusté individuellement pour chaque cible afin d'obtenir un taux d'erreur avoisinant 50%. Ainsi, au début du premier bloc expérimental, le niveau de contraste était situé exactement au centre de l'étendue disponible (valeur de 64 pour une étendue possible entre 0 et 128). Une fois les 20 premiers essais complétés, le contraste du masque était augmenté si la performance aux 10 essais précédents pour une cible était supérieure à 50% correct et diminué si celle-ci était inférieure (aucun ajustement n'était appliqué avant qu'une cible n'ait été présentée 10 fois). La taille initiale de ce changement était de 16 unités et cette valeur était réduite de moitié lorsqu'une réduction du contraste était

suivie d'une augmentation, ou vice-versa, jusqu'à un minimum de 1. L'état de l'algorithme contrôlant le contraste du masque de bruit blanc superposé à la cible était sauvegardé à la fin de chaque bloc et récupéré au début du suivant.

Les objets étaient partiellement révélés à travers douze d'ouvertures gaussiennes et étaient présentés pour une durée de 100ms. Les participants devaient dire le numéro associé au stimulus présenté et l'expérimentateur devait entrer les réponses des participants sur un clavier filaire. Un signal sonore était ensuite émis, les informant sur la validité de la réponse fournie. Une tonalité aiguë (1000 Hz, 100 ms) indiquait qu'ils avaient répondu correctement et une tonalité grave (300 Hz, 300 ms) qu'ils avaient fourni une réponse incorrecte. La feuille contenant les stimuli utilisés lors de la phase de familiarisation était disposée sous l'écran afin que les participants puissent s'y référer au besoin. (Voir figure 1)

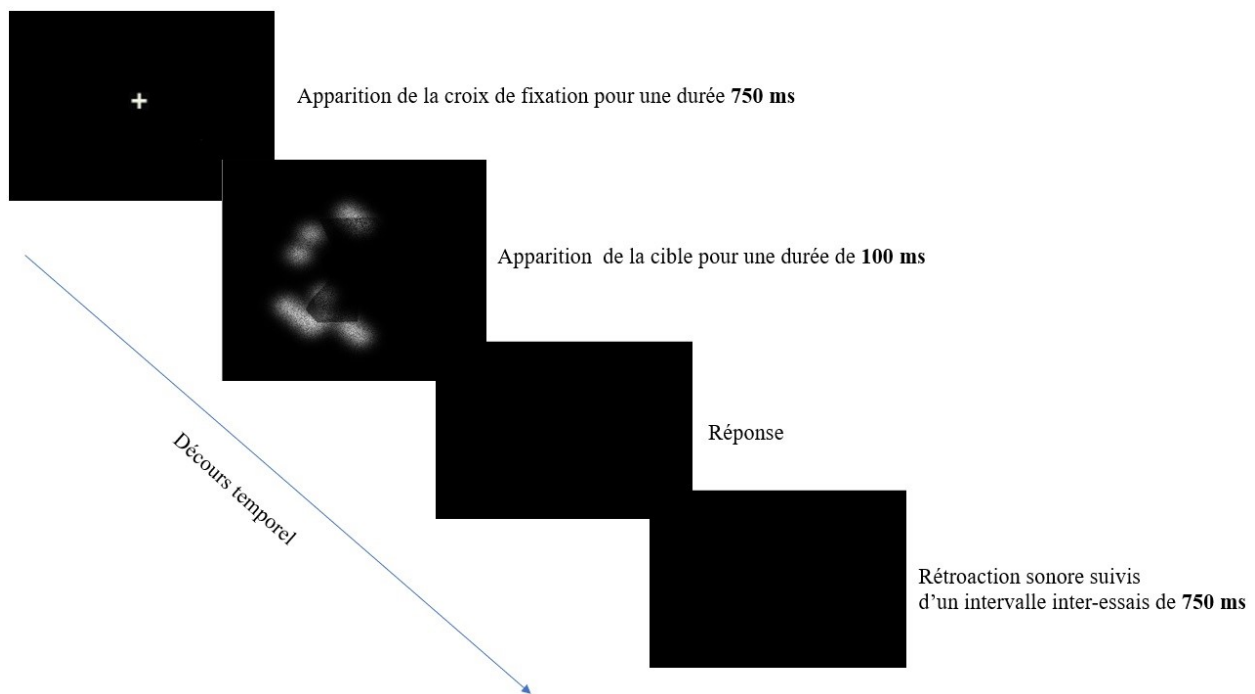


Figure 3 – Schéma illustrant le déroulement temporel de la tâche de reconnaissance effectuée par les participants humains.

### **La tâche de l'observateur idéal**

L'observateur idéal a effectué la même tâche de reconnaissance d'objets que les participants humains, dans laquelle il devait tenter d'établir une correspondance entre le stimulus présenté à un essai et une des cibles potentielles. À cette fin, des corrélations étaient calculées entre l'image du



stimulus et celle de chaque cible potentielle. Le coefficient de corrélation le plus élevé dictait la réponse fournie. Dans ce contexte, l'observateur idéal est utilisé pour déterminer la localisation de l'information contenue dans l'image qui permettait de discriminer entre les items cibles. Celui-ci fournit une valeur informative absolue, puisqu'aucune limite ou contrainte n'est imposée dans sa capacité à utiliser pleinement l'information disponible.

## **Traitement des données**

### **La construction d'images de classification brutes.**

La première étape de l'analyse des données consistait à construire des images de classifications individuelles (IC) brutes pour chacune des cibles et chacun des 12 participants. Les données utilisées à cette fin sont les masques ayant servi pour l'échantillonnage du stimulus et, de façon correspondante, les réponses correctes et les erreurs. Cette étape nous informe sur les parties de l'image qui se sont avérées efficaces à l'identification de chaque cible pour chacun des participants.

### **Le calcul de l'image de classification**

L'image de classification brute est calculée par une opération où la moyenne des masques associés aux erreurs est soustraite de la moyenne des masques associés aux réponses correctes.

### **La transformation des images de classification individuelles brutes en score Z**

Une fois les IC brutes calculées, celles-ci ont été transformées en scores Z. Cette opération consiste à soustraire des scores bruts une moyenne de référence représentative de l'ensemble de la distribution des données d'origine, pour ensuite diviser ce résultat par l'écart type de la distribution.

### **Le calcul de la moyenne et de l'écart-type de référence par la technique de « bootstrapping »**

La technique de « bootstrapping » consiste à produire une vaste collection d'IC aléatoires (1000 dans le cas présent) construites à partir d'une mise en correspondance aléatoire avec remplacement entre les localisations échantillonnées (c.-à-d.; les masques) et le caractère correct ou incorrect des

réponses obtenues. La moyenne de ces IC aléatoires et son écart-type sont ensuite utilisés pour la transformation des IC brutes en IC en scores Z.

### **Le moyennage des images de classification individuelles.**

Une fois les IC individuelles mises sur une échelle commune, soit celle des scores Z, leur moyenne fut calculée et l'IC de groupe moyenne pour chaque stimulus fût surimposée sur l'image de la cible correspondante.

### **Le test Pixel**

Avant d'effectuer le test Pixel (Chauvin, Worsley, Schyns, Arguin & Gosselin, 2005), nous avons procédé au lissage des IC avec un opérateur gaussien circulaire dont l'écart-type est de 0,31 degrés d'angle visuel. Le test Pixel fut ensuite appliqué sur l'image afin d'obtenir le seuil en score Z permettant d'identifier les régions de l'image de classification se distinguant significativement de 0. Dans le cas présent, un test bidirectionnel avec un seuil alpha global de 0,05 fut appliqué. Si une valeur donnée sur une image de classification est supérieure au seuil supérieur (i.e. celui supérieur à 0) du test Pixel, ceci indique que la zone de l'image de classification concernée contribue de manière significative à la capacité des participants à reconnaître la cible correctement. Au contraire, si une valeur est inférieure au seuil inférieur (i.e. celui inférieur à 0) du test Pixel, ceci indique que la zone de l'image de classification concernée contribue de manière significative à amener les participants à produire une réponse erronée.

### **Évaluation du nombre de participants partageant une région de l'image comme significativement supérieure / inférieure à 0.**

Afin d'évaluer dans quelle mesure les IC moyennes représentent adéquatement les résultats des participants individuels y ayant contribué, le nombre de participants partageant des régions de l'image de chaque cible significativement supérieures ou inférieures à zéro a été calculé. À cette fin, le test Pixel a été appliqué sur chaque image de classification de chaque participant après lissage avec le même opérateur que celui utilisé pour les images de classification moyennes. Une valeur de 1 a été attribuée à chaque point des images de classification significativement supérieur à 0 et -1 pour chaque point significativement inférieur à 0. Les points non-significativement différents de zéro ont reçu la valeur 0. Des IC cumulatives ont ensuite été calculées en additionnant,

à travers tous les participants, les valeurs attribuées à chaque point de la surface de chaque cible. La valeur maximale sur ces IC cumulatives correspond donc au nombre de participants, soit 12, et le minimum est de -12. Enfin, des corrélations ont été calculées afin de quantifier la correspondance entre les IC cumulatives et les IC moyennes.

### **Corrélation des images de classification des observateurs humains vs observateur idéal :**

L'observateur idéal sert ici à déterminer les localisations de l'image offrant une information utile pour discriminer entre les cibles. Celui-ci a réalisé la même tâche et a complété le même nombre d'essais que les observateurs humains. Des images de classification de l'observateur idéal ont été construites pour chaque cible et celles-ci ont été soumises au test Pixel, suivant la même méthode que pour le traitement des données des observateurs humains. Ces IC ont ensuite été comparées aux IC moyennes des observateurs humains par le biais de corrélations.

### **Les analyses de « zone unique »**

Le fait que les stimuli aient toujours été présentés à la même localisation à chaque essai soulève la possibilité de l'utilisation, par les participants, de la localisation d'une aperture révélant la cible sur l'écran qui aurait pu agir comme un indicateur de l'identité de la cible. Afin d'avoir une indication sur la possibilité qu'une telle stratégie ait été employée par les participants, la surface occupée par chaque cible fut calculée et les surfaces occupées par toutes les autres cibles en furent soustraites. L'échelle de couleur (voir Fig. 4) révèle les zones appartenant plus ou moins uniquement à la cible. Des corrélations entre les images obtenues selon cette méthode et les images de classification moyennes furent effectuées afin de quantifier cette correspondance.

# Résultats

## Performance et images de classification des participants humains

Les taux moyens de réponses correctes pour chaque cible sont présentés au Tableau 1. Ceux-ci sont très proches de 50%, tel que prévu à travers l'algorithme contrôlant le niveau de contraste du masque de bruit blanc superposé à la cible. Une exception concerne l'objet no. 7, pour lequel le taux de réponses correctes est de 60%. Ce taux demeure cependant tout à fait acceptable pour le calcul des images de classification, qui demande des nombres d'essais suffisamment élevés autant pour les réponses correctes que pour les erreurs. Les niveaux de contraste moyens du bruit blanc surimposé à chaque cible sont indiqués au Tableau 2. Ceux-ci sont très bas, hormis pour l'objet no. 3, qui a demandé l'application d'un niveau de bruit nettement plus élevé que les autres objets afin de maintenir la performance des participants autour de 50% de réponses correctes. Ceci indique que cet objet était nettement plus facile que les autres à reconnaître.

**Tableau 1 – Taux moyen de réponses correctes (%) pour les observateurs humains pour chaque objet cible lors des essais expérimentaux.**

Images cibles	1	2	3	4	5	6	7	8
Taux (%) moyen de réponses correctes	52,24	51,68	50,36	50,57	50,78	50,79	60,16	50,90

**Tableau 2 – Niveau de bruit moyen du masque de bruit blanc surimposé à l'image de chaque cible aux essais expérimentaux réalisés par les observateurs humains.**

Images cibles	1	2	3	4	5	6	7	8
Niveau de bruit moyen	1,74	1,91	72,38	6,72	7,37	1,74	1,80	1,79

Les IC moyennes sur lesquelles le test Pixel a été appliqué sont présentées sur la Fig. 4. Chaque IC illustre l'efficacité avec laquelle les participants ont utilisé l'information disponible à chaque point de l'image pour identifier la cible. Les régions des IC qui sont achromatiques présentent des

scores Z ne se distinguant pas significativement de 0. Autrement dit, leur score Z ne dépasse pas les seuils fixés par le test Pixel. Ces seuils sont indiqués au Tableau 3 et sur la légende de chaque IC. L'étendue de la légende de chaque IC colorée en noir correspond à l'intervalle des scores Z non-significatifs. L'étendue de la légende au-dessus de la portion en noir réfère à des scores Z positifs et significativement différents de 0. Les régions de l'IC avec ces couleurs apportent une contribution favorable à l'identification correcte de la cible lorsqu'elles sont visibles (i.e. effet facilitateur). Au contraire, l'étendue de la légende en-dessous de la portion en noir réfère à des scores Z significativement inférieurs à 0, indiquant ainsi un impact négatif sur les performances pour les régions de l'IC portant ces couleurs (i.e. effet inhibiteur).

**Tableau 3 – Seuil du test Pixel pour chaque image de classification moyenne pour les observateurs humains.**

Images cibles	1	2	3	4	5	6	7	8
Seuil Pixel	4,47	4,68	4,02	4,62	4,19	4,51	1,96	1,96

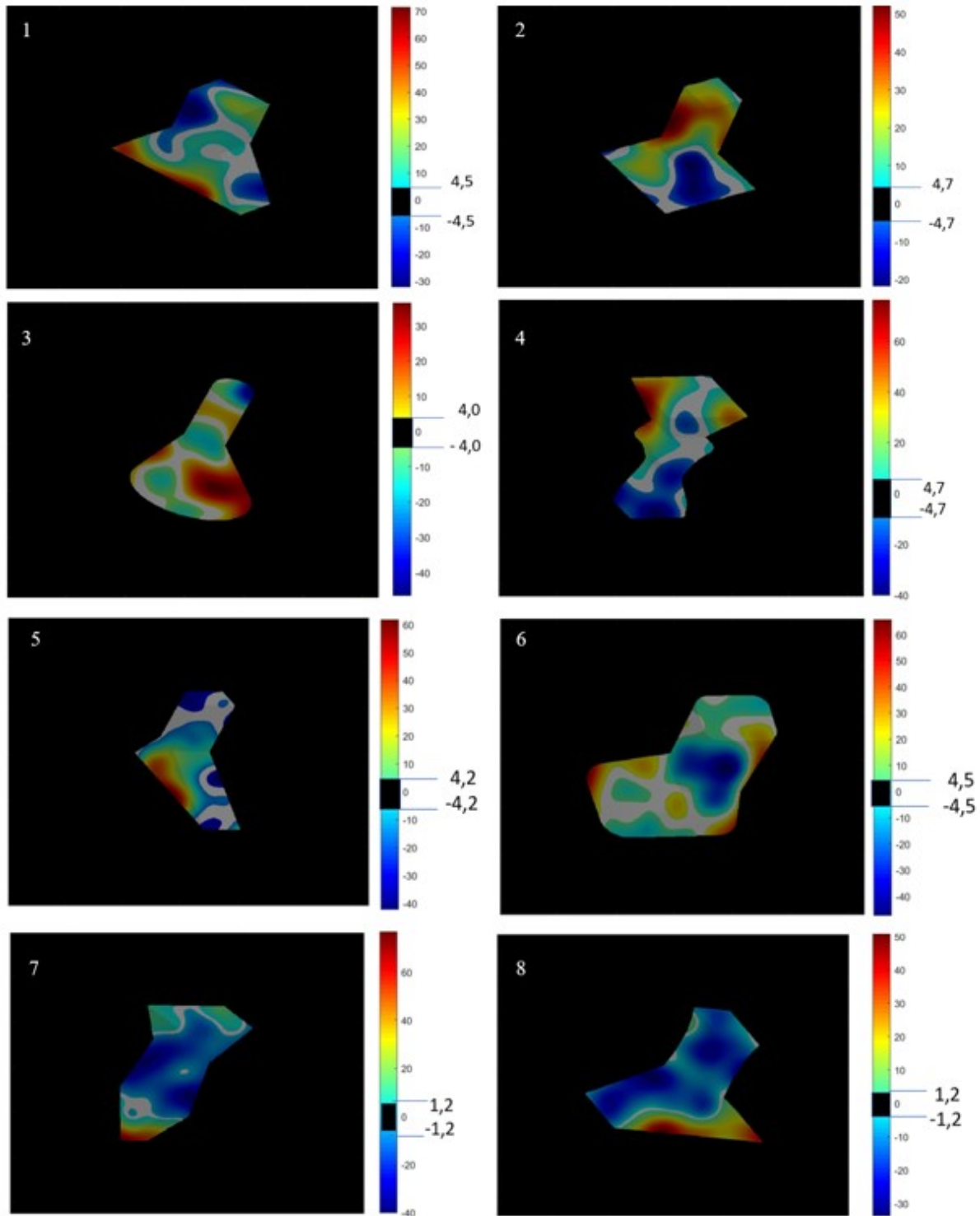


Fig 4. L'échelle de couleur fait référence à l'efficacité de traitement (scores Z). Les conventions d'illustration sont décrites dans le texte

Figure 4 – Images de classification moyennes pour les participants humains.

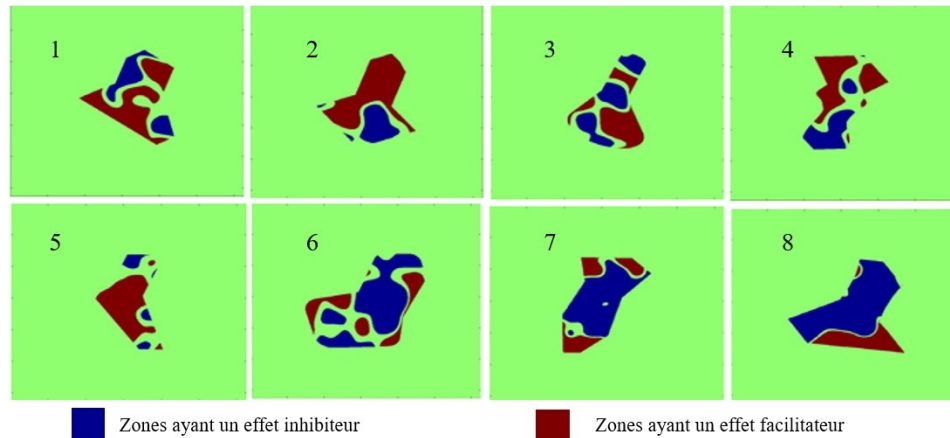


Figure 5 – Images de classifications simplifiées illustrant les zones ayant un effet facilitateur ou inhibiteur sur la performance de reconnaissance de chaque cible par les observateurs humains.

### Les images de classification cumulatives

Comme les seuils de signification obtenus par le test Pixel ne prennent pas en considération la variabilité inter-sujets, il semble pertinent de vérifier si les phénomènes observés dans les IC moyennes sont représentatifs des IC individuelles des observateurs constituant l'échantillon. À cette fin, une série d'IC cumulatives fut générée (Fig. 5). Les IC cumulatives ont été construites de la manière suivante : 1- le test Pixel a été appliqué sur chaque IC individuelle; 2- chaque point appartenant aux régions de l'IC ayant produit un effet facilitateur s'est vu attribuer une valeur de 1 alors que chaque point associé à un effet inhibiteur a reçu une valeur de -1; les zones non-significatives ont une valeur de 0; 3- les images de classification de tous les participants pour chaque objet ont été additionnées. Les couleurs chaudes dans les IC cumulatives représentent les régions ayant eu un effet facilitateur pour un nombre important de participants. Ainsi, le rouge le plus foncé représente un nombre élevé de participants présentant un effet facilitateur pour la région concernée (maximum possible de +12) et le bleu foncé indique une région de l'image ayant eu un effet inhibiteur chez un nombre élevé d'observateurs (minimum possible de -12).

Les IC moyennes (Fig. 4) furent ensuite comparées aux IC cumulatives (Fig. 6) en calculant le niveau de corrélation des images correspondantes. Ces coefficients de corrélation sont rapportés dans le Tableau 4.

Les corrélations obtenues sont très élevées, celles-ci se situant entre 0,66 et 0,84. Ceci démontre donc que les phénomènes observés dans les IC moyennes constituent une représentation valide de ceux concernant les individus constituant le groupe.

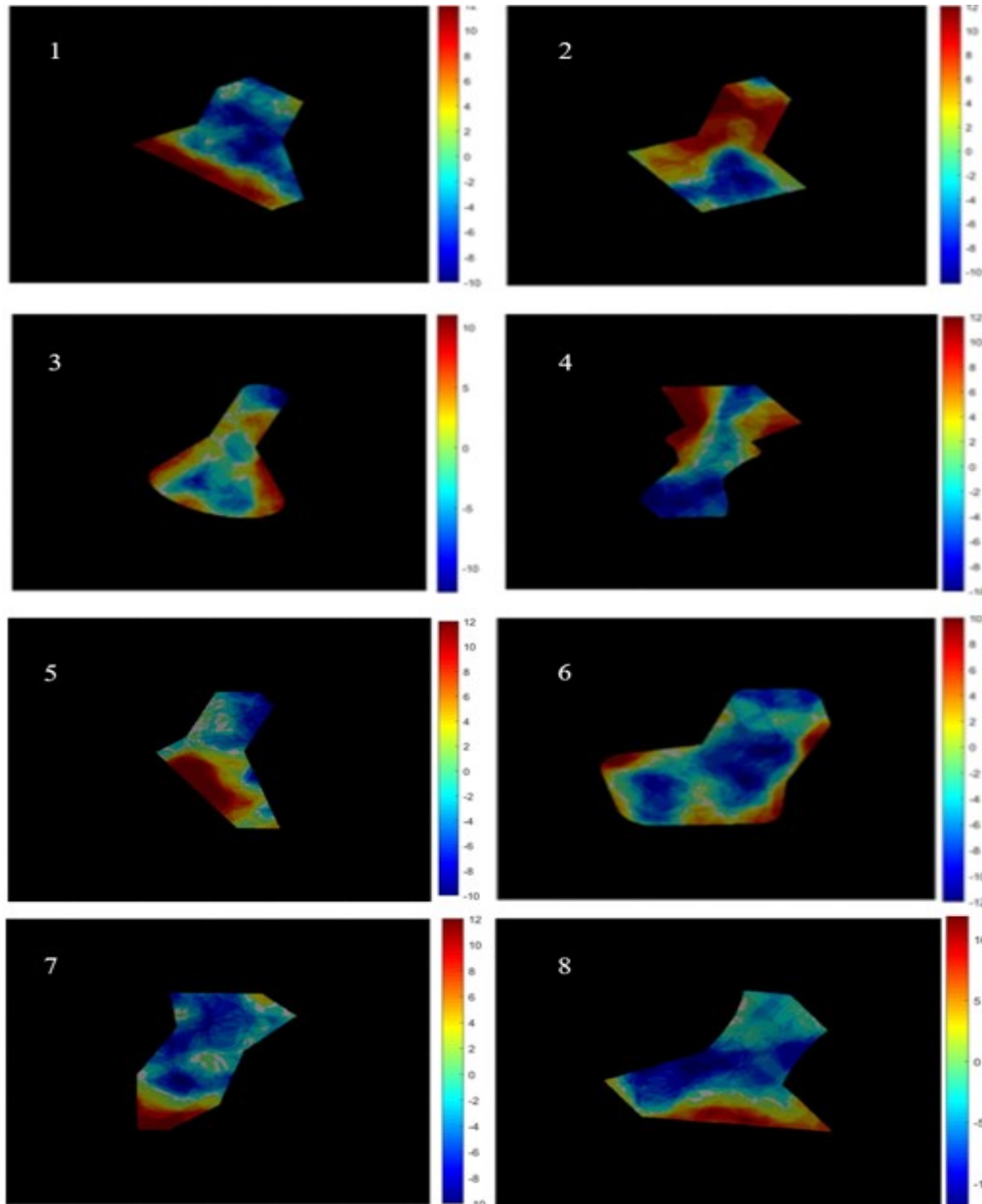


Fig 6. L'échelle de couleur réfère à l'aspect facilitateur ou inhibiteur de chaque point de l'image, calculé en nombre de participants. Les autres conventions d'illustration sont décrites dans le texte.

Figure 6 – Images de classification cumulatives des participants humains pour chaque cible.



**Tableau 4 – Valeurs du coefficient de corrélation (r) entre les images de classification moyenne et les images de classification cumulatives pour chacune des cibles.**

Images cibles	1	2	3	4	5	6	7	8
Coefficient de corrélation (r)	0,67	0,79	0,77	0,81	0,71	0,74	0,83	0,75

### **Les zones uniques**

Tel qu'indiqué dans la description de la méthodologie, les stimuli présentés aux participants étaient toujours centrés sur le point de fixation. Autrement dit, à chaque occasion où une cible était présentée, celle-ci apparaissait exactement à la même localisation sur l'écran. En fin de cueillette de données, nous avons réalisé que cette situation permet potentiellement aux participants d'appliquer une stratégie par laquelle la simple visibilité du stimulus à une localisation particulière sur l'écran pouvait être informative quant à l'identité de la cible (stratégie de localisation diagnostique). La validité d'une telle stratégie repose principalement sur le fait que les contours externes de la cible sont localisés à des endroits différents sur l'écran selon l'objet qui est présenté.

Afin d'examiner si les participants ont appliqué une telle stratégie pour réaliser la tâche, une analyse supplémentaire a été effectuée afin de révéler les zones de chevauchement de l'image à travers les cibles ainsi que les zones qui leurs sont uniques. Pour chaque cible, une valeur de 1 a été attribuée à chaque point de l'image occupée par le stimulus. Ensuite, une valeur de 0,1 a été soustraite de chaque localisation occupée à la fois par l'image de la cible et celle d'un autre objet, et ce pour tous les autres objets. Le résultat de cette opération est illustré à la Fig.7. L'échelle de couleurs reflète une étendue de valeurs numériques entre 0 et 1. Une valeur de 1 correspond à une zone qui est unique à la cible et moins la valeur est élevée (jusqu'à un minimum de 0,3) plus elle correspond à une région qui est partagée avec d'autres cibles. Autrement dit, le code de couleur reflète la valeur diagnostique d'une région de l'image pour l'identité de la cible.

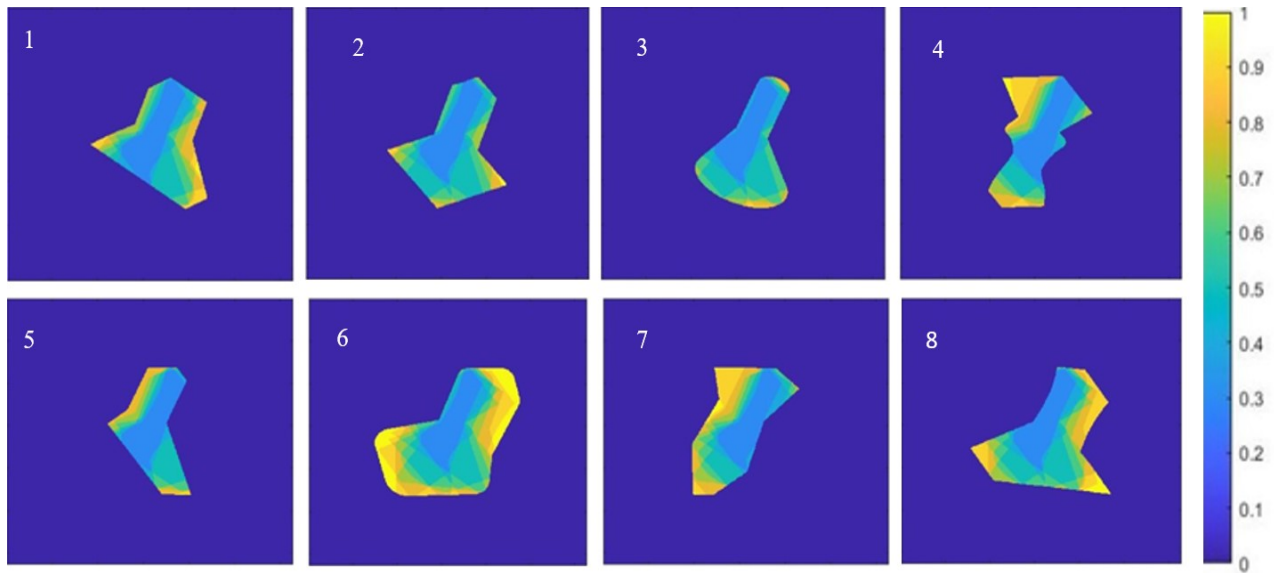


Fig 7. Les conventions d’illustration sont décrites dans le texte.

Figure 7 – Illustration représentant les zones uniques pour chaque cible.

Advenant que les participants aient utilisé une stratégie de localisation diagnostique, il devrait exister une corrélation forte entre les illustrations de la Fig. 10 et les IC moyennes décrites plus haut. Les valeurs des coefficients de corrélation quantifiant cette correspondance sont indiquées au Tableau 5 ci-dessous. Ces corrélations sont toutes négatives, démontrant ainsi que les participants n’ont pas appliqué une stratégie de localisation diagnostique pour réaliser la tâche.

**Tableau 5 – Valeurs des coefficients de corrélation entre les zones uniques et l’IC moyenne pour chaque cible.**

Images cibles	1	2	3	4	5	6	7	8
Coefficient de corrélation	-0,09	-0,56	-0,50	-0,34	-0,17	-0,14	-0,20	-0,34

## Performance et images de classification de l’observateur idéal

Comme pour les participants humains, le taux moyen de réponses correctes visé pour chaque cible devait avoisiner 50 %. Cependant, les résultats de l’observateur idéal, lorsque testé sous les mêmes conditions que les observateurs humains, ne se conforment pas à cet objectif. Ainsi, les taux de

réponses correctes étaient extrêmement bas pour une majorité de cibles (Tableau 5) alors que le contraste du bruit blanc surimposé aux images des cibles était minimal (Tableau 6). Étonnamment donc, l'observateur idéal s'est avéré incapable d'atteindre un niveau de performance acceptable alors qu'il était soumis aux mêmes conditions de stimulation que les participants humains. Il s'agit d'un résultat inusité qui n'avait pas été anticipé. La discussion abordera notre hypothèse quant à l'explication de cette observation.

**Tableau 6 – Taux moyen de réponses correctes (%) et niveaux moyen de bruit blanc pour l'observateur idéal avec un niveau de contraste à 20 % des images des cibles.**

Images cibles	1	2	3	4	5	6	7	8
Taux (%) moyen de réponses correctes	2,32	17,68	50,89	36,61	32,32	1,25	11,43	15,36
Niveau de bruit moyen	1,74	1,91	72,38	6,72	7,37	1,74	1,79	1,79

Après quelques tâtonnements, nous avons constaté que l'observateur idéal arrive à atteindre les niveaux de performance souhaités lorsque le niveau de contraste des images des cibles était augmenté à 50% (rappel : le contraste des images était de 20% pour les participants humains). Les taux de réponses correctes de l'observateur idéal sous ces conditions sont indiqués au Tableau 7. Les niveaux moyens de contraste du bruit blanc surimposé aux cibles pour atteindre ces performances sont donnés au Tableau 8.

**Tableau 7 – Taux moyen de réponses correctes (%) pour l'observateur idéal avec un niveau de contraste de 50% des images des cibles.**

Images cibles	1	2	3	4	5	6	7	8
Taux (%) moyen de réponses correctes	47,5	48,93	50,71	51,07	50,0	50,18	50,36	50,36

**Tableau 8 – Niveaux moyens de contraste du bruit blanc pour l’observateur idéal avec un niveau de contraste de 50% des images des cibles.**

Images cibles	1	2	3	4	5	6	7	8
Niveau de bruit moyen	74,84	78,83	05,40	89,32	84,60	75,44	82,01	83,96

Les IC de l’observateur idéal (Fig. 8) furent calculées suivant les mêmes procédures que pour les participants humains et les conventions d’illustration sont les mêmes qu’à la Fig. 4. Les seuils établis par le test Pixel pour chaque cible sont indiqués au Tableau 9.

**Tableau 9 – Seuil du test Pixel pour chaque image de classification pour l’observateur idéal.**

Images cibles	1	2	3	4	5	6	7	8
Valeur du seuil Pixel	1,96	1,96	4,61	4,58	4,51	4,56	4,40	3,94

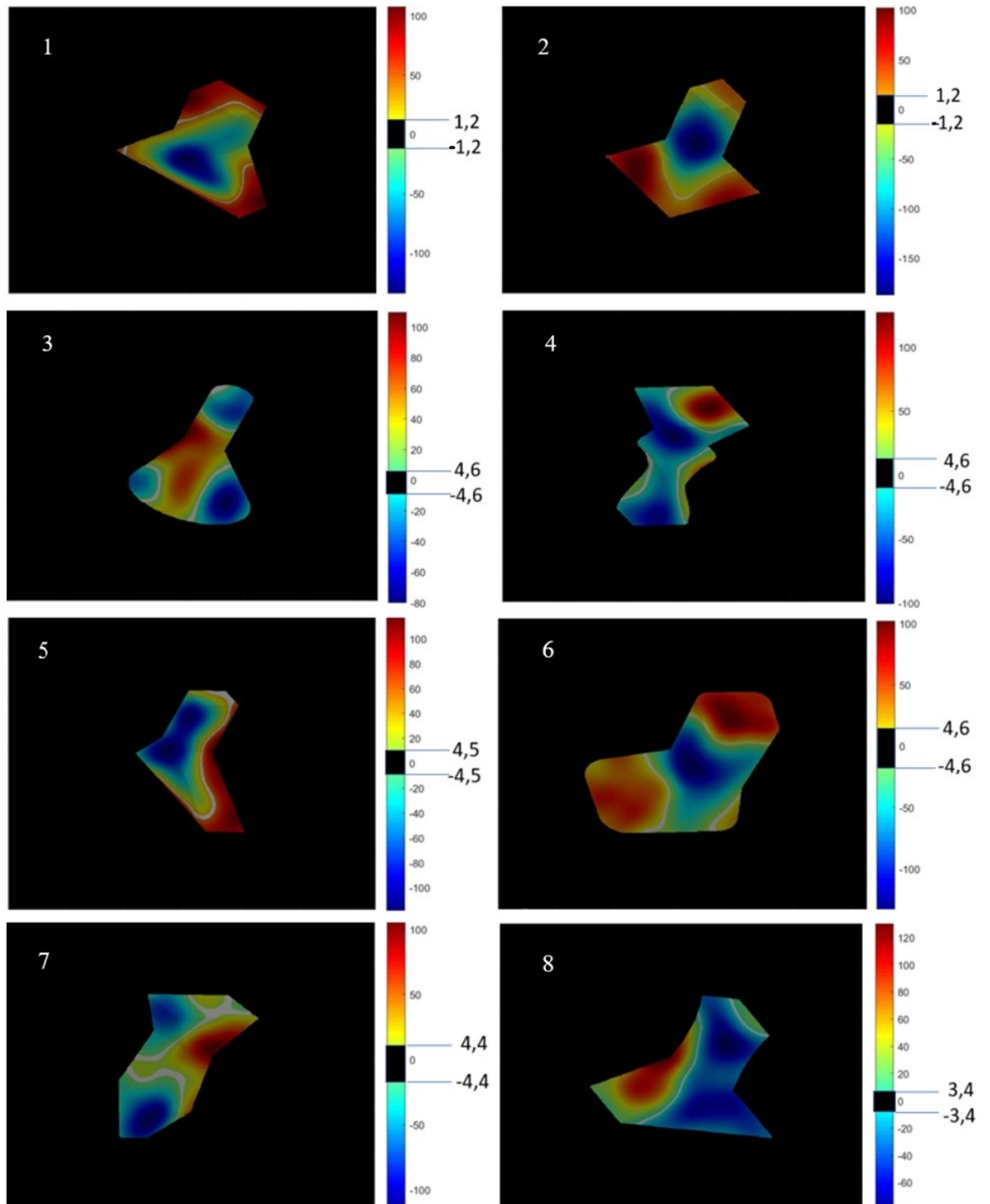


Fig 8. L'échelle de couleur fait référence à l'efficacité de traitement (scores Z). Les conventions d'illustration sont les mêmes qu'à la Fig 4.

Figure 8 – Images de classification moyennes pour l'observateur idéal.

Les images de classifications simplifiées de l'observateur idéal sont illustrées à la Fig. 9. Celles-ci distinguent les régions de l'image des cibles associées à des effets significatifs facilitateurs (en rouge) ou inhibiteurs (en bleu). Les minces bandes de même couleur que le fond sur ces figures correspondent aux régions de l'image n'ayant aucun impact significatif sur la capacité de l'observateur idéal à reconnaître la cible.

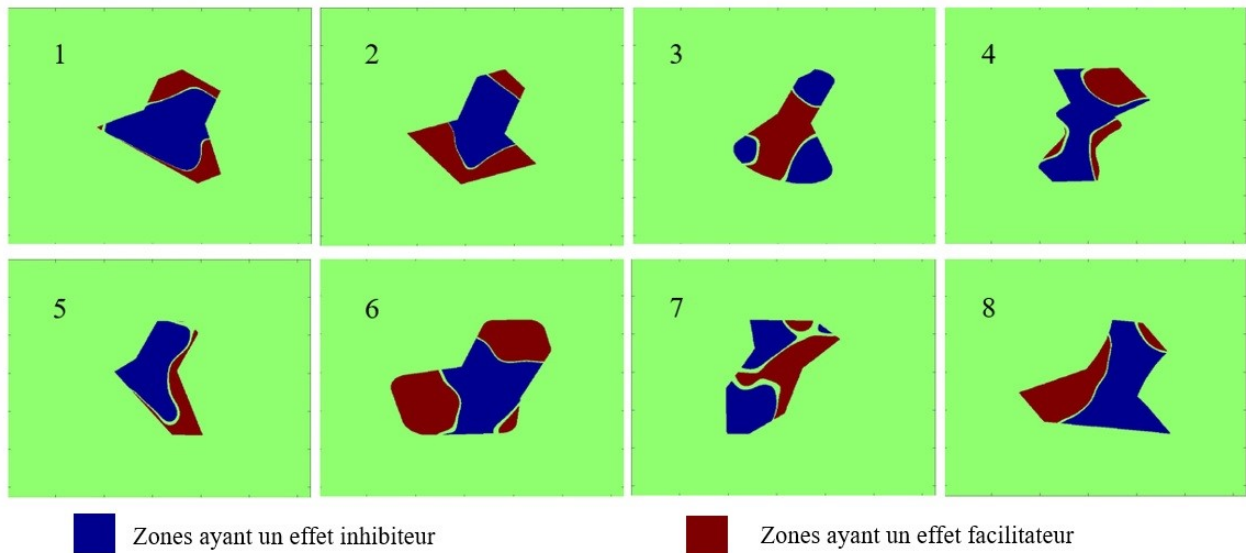


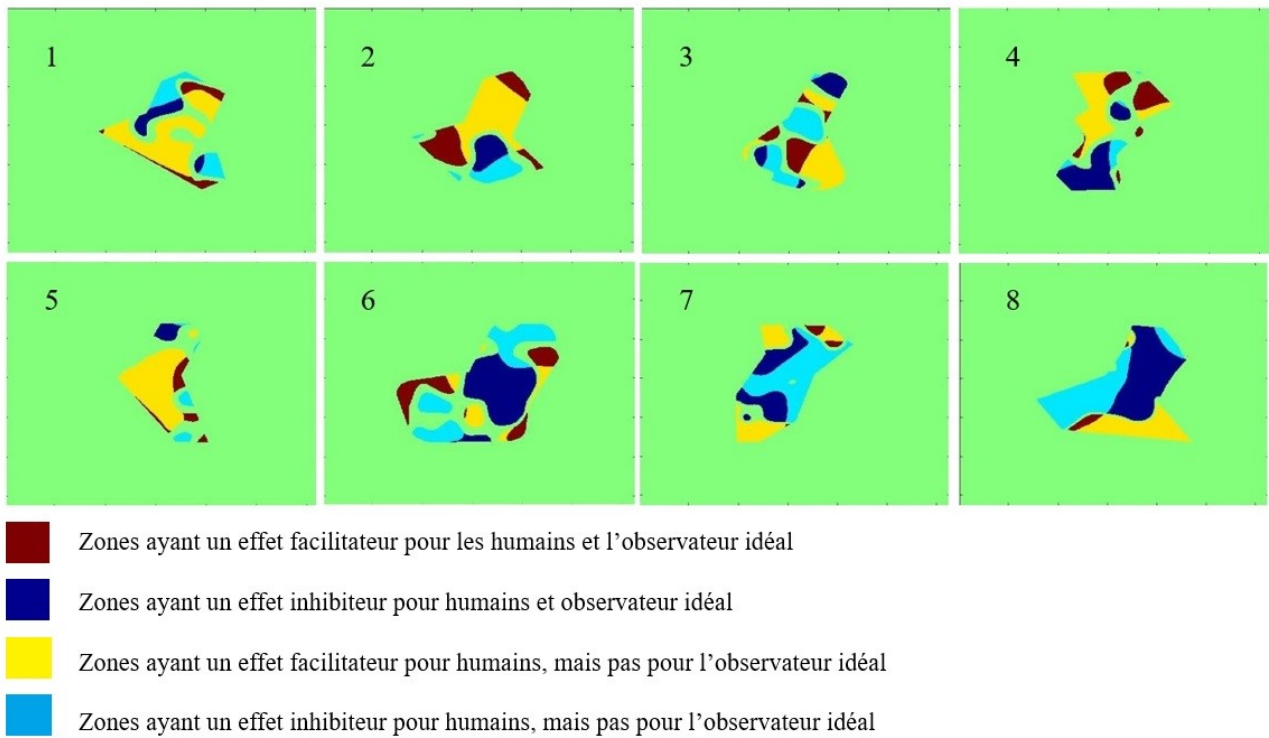
Figure 9 – Images de classifications simplifiées illustrant les zones ayant un effet facilitateur ou inhibiteur sur la performance de reconnaissance de chaque cible par l'observateur idéal.

## Observateurs humains vs observateur idéal

Des corrélations ont été calculées entre les IC moyennes obtenues par les participants humains et les IC correspondantes de l'observateur idéal (Tableau 10). Les coefficients de corrélation sont très faibles et deux d'entre eux sont négatifs, indiquant ainsi l'absence d'une correspondance systématique entre les deux ensembles d'IC. Ceci est également confirmé par les images de contraste illustrées à la Fig. 10. Sur ces dernières, les zones en rouge y représentent les régions sur l'image des cibles ayant un effet facilitateur à la fois pour les humains et l'observateur idéal. Les zones bleu foncé celles ayant un effet inhibiteur à la fois pour les humains et l'observateur idéal. Les zones jaunes représentent celles ayant un effet facilitateur et les zones bleu pâle celles ayant un effet inhibiteur uniquement pour les humains.

**Tableau 10 – Valeurs du coefficient de corrélation (r) entre les images de classification moyennes des participants humains et celles de l’observateur idéal, pour chacune des cibles.**

Images cibles	1	2	3	4	5	6	7	8
Coefficient de corrélation (r)	-0,22	0,38	0,46	0,28	-0,09	0,29	0,15	0,33



*Fig 10. Les conventions d’illustration sont décrites précédemment dans le texte.*

**Figure 10 – Images de contraste entre les IC des participants humains et de l’observateur idéal.**

## Discussion

Cette étude est parvenue à déterminer les régions de l'image utilisée par les participants afin de reconnaître chacun des 8 objets non familiers faits de deux composantes. En effet, les zones ayant un effet facilitateur et celles ayant un effet inhibiteur sur les performances humaines furent identifiées ainsi que celles d'un observateur idéal régulier. La technique des bulles a effectivement révélé certaines régions de chaque cible comme ayant été plus ou moins utiles que d'autres et a permis de démontrer que les différentes parties de l'image ne contribuent pas de manière égale à la reconnaissance de l'objet y étant illustré.

Les IC moyennes des participants humains ont été comparées aux IC cumulatives, ces dernières représentant un cumul des régions de l'image ayant eu un impact significatif sur l'efficacité de traitement des participants individuels. La forte convergence des deux types d'images de classification indique que les IC moyennes constituent une représentation valide du traitement réalisé par les participants individuels constituant le groupe. Le reste de la discussion se focalisera donc, sur les résultats moyens du groupe de participants humains en prenant pour acquis que ceux-ci s'appliqueront de manière valide aux individus.

Les IC moyennes ont également été comparées aux images représentant les « zones uniques » des images des cibles. Ces zones uniques reflètent le degré de chevauchement entre les surfaces de chaque image cible avec celles des autres objets utilisés dans le contexte de la présente expérience. L'objectif visé par cette analyse était d'examiner dans quelle mesure les IC auraient pu être déterminées par l'application d'une stratégie de « localisation diagnostique » par les participants. Une telle stratégie devrait être basée sur la présence dans la stimulation de certaines zones diagnostiques de l'image qui sont occupées par un objet et donc, plus une zone de l'image d'une cible est « unique », plus sa valeur diagnostique devrait être élevée. L'analyse indique des corrélations qui sont toutes négatives, démontrant ainsi qu'une stratégie de « localisation diagnostique » n'est pas responsable des IC observées chez les participants humains. Il semble donc possible de conclure que les IC moyennes reflètent l'application des mécanismes de reconnaissance d'objets usuels plutôt qu'un artéfact lié à l'application d'une stratégie motivée exclusivement par le contexte particulier de la présente expérience.



## **Analyse comparative des IC des humains avec celles de l'observateur idéal**

Les IC moyennes ont aussi été comparées aux IC obtenues par l'observateur idéal. Ces dernières sont censées représenter la quantité d'information pertinente à la discrimination entre les cibles pour chaque région de la surface occupées par ces dernières. Tel qu'indiqué dans la section « Résultats », la technique des bulles a effectivement révélé certaines régions de chaque cible comme ayant été plus ou moins utiles que d'autres à la reconnaissance

L'observateur idéal s'est avéré être un très mauvais prédicteur des IC moyennes des participants humains comme l'indique les faibles corrélations (proches de zéro ou négatives) observées. Il semble donc n'exister aucune correspondance systématique entre les zones diagnostiques utilisées par les humains et l'observateur idéal pour reconnaître les cibles (voir Tableau 10). Il était attendu que les informations diagnostiques révélées par les IC des participants humains correspondent à un sous-ensemble des régions efficaces de l'image pour l'observateur idéal. Quoique ceci semble vrai pour certains objets, la règle ne semble pas s'appliquer de manière évidente pour d'autres (voir Fig. 10). Une partie des régions diagnostiques ayant servi à l'identification correcte de la cible 2 correspond à des régions présentant également un effet facilitateur significatif pour l'observateur idéal. Les objets 4 et 6 constituent également d'autres bons exemples à cet égard. Cependant, le chevauchement des zones présentant un effet facilitateur significatif pour les observateurs humains et l'observateur idéal est minimal pour d'autres objets. C'est notamment le cas des cibles 1, 5, 7 et 8 (Voir Fig.10).

Un autre aspect concernant les performance de l'observateur idéal mérite d'être abordé ici. Ce dernier n'était pas en mesure d'effectuer la tâche dans les mêmes conditions d'exposition auxquelles les participants humains ont performé au niveau souhaité (i.e. taux approximatif de 50 % de réponses correctes). Afin que l'observateur idéal atteigne le niveau de performance souhaité, le contraste des images des cibles dû être augmenté de 20% (tel que pour les observateurs humains) à 50%.

Cette situation était complètement inattendue compte tenu que l'observateur idéal, contrairement aux participants humains, peut utiliser toute l'information discriminante disponible contenue dans l'image, sans aucune limitation. Au contraire les humains sont confrontés à des facteurs tels

l'acuité limitée qui diminue avec l'excentricité, la courbe de sensibilité au contraste, etc. En vertu des limitations humaines, il était prévu que l'observateur idéal arrive facilement à faire la tâche dans les mêmes conditions que celles utilisées pour les humains, avec un taux de réponses correctes d'au moins 50% et des niveaux de contraste du bruit blanc surimposé à l'image de la cible égaux ou supérieurs à ceux requis pour les observateurs humains.

La supériorité de la capacité de traitement des humains constatée dans les résultats semble pouvoir s'expliquer par un phénomène de complétion amodale, qui est connu pour s'appliquer dans le fonctionnement visuel humain lorsque confronté à de l'occlusion, comme dans la présente expérience. Par opposition, aucune capacité de complétion amodale n'était disponible à l'observateur idéal.

La complétion amodale fait référence à la complétion d'un objet en l'absence de stimulation visuelle directe lorsque ce dernier n'est que partiellement visible en raison d'occlusion. Ce type de complétion est qualifié d'amodale car les parties cachées de l'objet ne peuvent être vérifiées par aucune modalité sensorielle.

L'occlusion des objets dans notre environnement constitue un défi important auquel le système visuel humain est continuellement confronté. Des études d'imagerie cérébrale fonctionnelle par résonance magnétique (fMRI) indiquent que la partie occluse d'un objet est représentée topographiquement dans les aires V1 et V2 (Ban & al., 2013). Cela suggère que les neurones des aires V1 et V2, en plus d'être en mesure de traiter les contours visibles, pourraient également représenter les parties d'un objet qui sont invisibles pour cause d'occlusion. La nature automatique de la complétion amodale est appuyée entre autres par des résultats provenant d'études comportementales indiquant qu'elle se déroule très rapidement dans le traitement visuel. Ainsi, dans le cas de stimulations picturales, elle surviendrait entre 200 et 400 ms après l'apparition d'un stimulus (Bruno, Bertamini, & Domini 1997; Sekuler & Palmer, 1992) alors qu'elle serait disponible dès 100 ms dans le cas d'une stimulation stéréoscopique (Bruno & al., 1997).

Nous proposons ici que la performance déficiente de l'observateur idéal sous des conditions permettant aux participants humains de fonctionner adéquatement puisse s'expliquer par un phénomène de complétion amodale. Spécifiquement, nous proposons que les observateurs humains étaient en mesure d'inférer correctement des parties des cibles qui étaient cachées par le masque d'occlusion, ce qui leur permettait ainsi d'augmenter l'information disponible sur la base

de laquelle la cible était identifiée. Au contraire, l'observateur idéal étant incapable de ce type d'inférence, se retrouvait ainsi avec une quantité d'information réduite relativement à celle disponible aux participants humains, expliquant ainsi ses capacités perceptives inférieures.

Dans le reste de la discussion, nous allons analyser d'une manière relativement détaillée les propriétés des zones diagnostiques de l'image des cibles afin de les caractériser au mieux possible et d'examiner dans quelle mesure les attributs impliqués sont ou non compatibles avec ceux proposés par les différentes théories structurales passées en revue dans l'Introduction.

## **Caractérisation des zones diagnostiques chez l'humain**

### **Attributs de bas niveau : contours, concavités/convexités et intersections**

La figure 11 représente les IC moyennes des participants humains, pour les cibles 1, 2, 4, 7 et 8. Pour chaque objet, parmi les zones présentant une haute valeur diagnostique (en rouge ou orange), y figure au moins une zone présentant un haut degré de convexité causée par une intersection de contours (zones encerclées en rouge). Par exemple, dans le cas de l'objet 7, soit celui ayant été le mieux reconnu en moyenne par les participants, la pointe à sa base s'est révélée très importante pour son identification. Des maxima de courbure causés par l'intersection de contours externes ou internes ont également eu une importance cruciale pour l'identification des objets numéro 1, 2, 4 et 8. Ces observations sont en accord avec l'affirmation que les maxima de courbure sont utiles à la caractérisation des éléments constitutifs d'une image (Bertramini & Wagemans, 2013). Ajoutons également que dans le cas de l'objet numéro 4, les régions les plus diagnostiques pour l'identification de l'objet comprennent également une portion du contour externe présentant une concavité marquée (i.e. minima de courbure négative), qui est postulé par certains auteurs comme une information déterminante pour la reconnaissance d'objet (Breakspears, Brammer, Das & William, 2004; Smith, Singh, Williams & Greenlee, 2001; Uyar, Shomstein, Greenberg & Behrmann 2016; Breakspear, Brammer & Robinson, 2003).

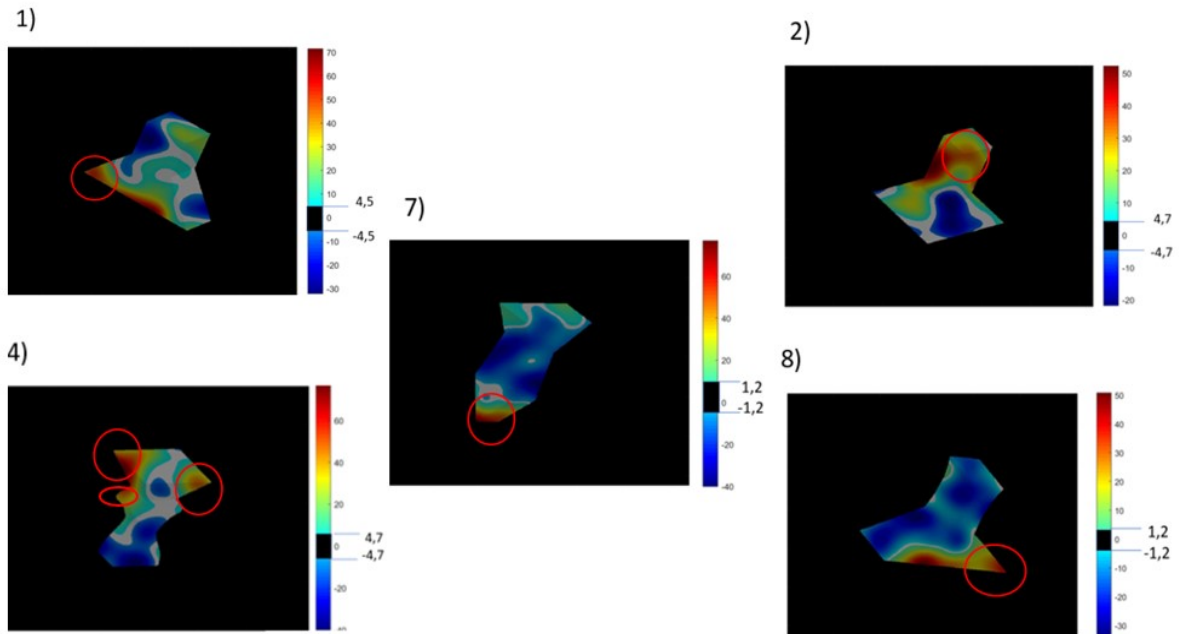


Fig 11. L'échelle de couleur fait référence à l'efficacité de traitement (scores Z). Les conventions d'illustration sont les mêmes qu'à la Fig 4.

Figure 11 – Images de classification moyennes des cibles 1, 2, 4, 7 et 8.

Il s'avère toutefois important de souligner que d'autres convexités causées par l'intersection de contours ou d'autres concavités marquées sont présentes sur les objets illustrés à la Fig. 11 ainsi que sur presque tous les autres objets de l'ensemble utilisé, mais que celles-ci n'ont pas contribué favorablement à l'identification des cibles. Cependant, comme on peut le constater à la Fig. 12, des maxima de convexité le long d'un contour continu constituent l'essentiel des zones de haute diagnosticité pour l'identification de la cible numéro 6.

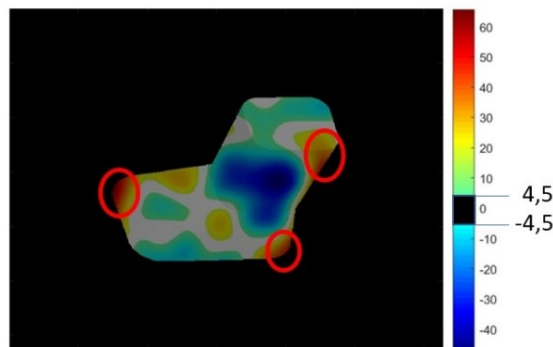


Fig 12. L'échelle de couleur fait référence à l'efficacité de traitement (scores Z). Les conventions d'illustration sont les mêmes qu'à la Fig 4.

Figure 12 – Image de classification moyenne de la cible 6.

Dans le cas des objets 1, 2, 5 et 8, un autre attribut ayant apporté une contribution de grande importance à leur identification consiste en un segment de contour externe droit (i.e. sans courbure) qui est suffisamment éloigné de toute convexité ou concavité de contour pour être considéré ne pas en faire partie (Fig. 13). Cependant, nous retrouvons une multitude d'autres contours droits dans notre ensemble de stimuli qui n'ont eu aucun apport facilitateur pour la réalisation de la tâche.

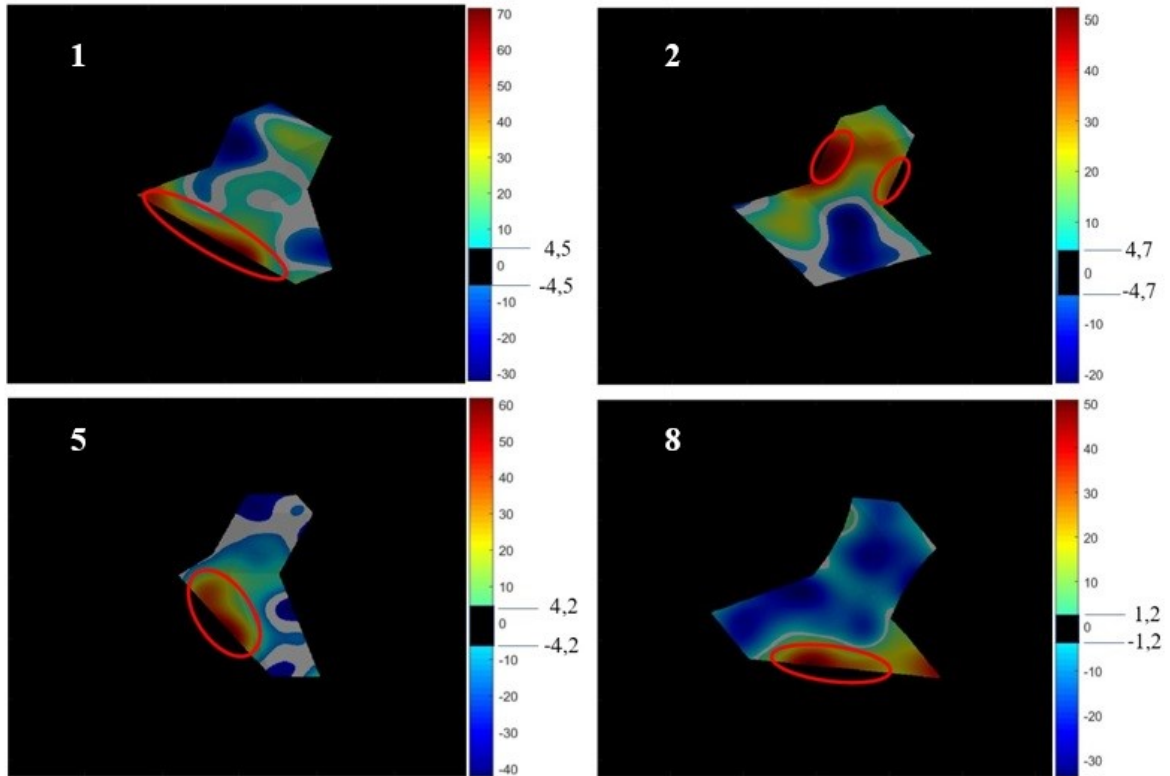


Fig 13. Les zones de couleur chaude située à l'intérieur des cercles rouges sont celles employées pour identifier ces cibles par le groupe d'humain normal. Les conventions d'illustration sont les mêmes qu'à la Fig 4.

Figure 13 – Images de classification moyennes des cibles 1, 2, 5 et 8 illustrant les zones facilitatrices.

Enfin, la cible numéro 3, constitue un cas unique parmi notre ensemble d'objets et dont l'identification semble reposer largement sur la courbure de la surface de sa partie inférieure, qui est située à l'intérieur du cercle rouge tel qu'illustré sur la Figure 14.

3)

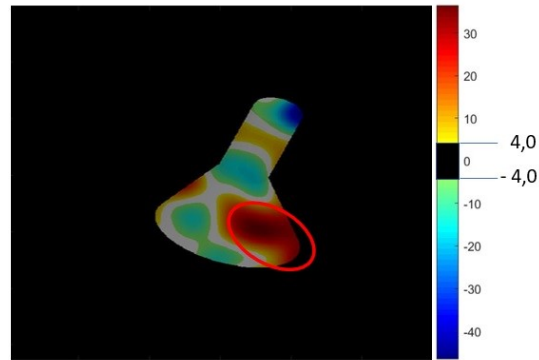


Fig 14. L'échelle de couleur fait référence à l'efficacité de traitement (scores Z). Les conventions d'illustration sont les mêmes qu'à la Fig 4.

Figure 14 – Image de classification moyenne de la cible 3.

En résumé, nous constatons que des attributs élémentaires hétérogènes caractérisent les régions diagnostiques des cibles utilisées dans la présente étude. Cependant, pour tous les attributs considérés, leur présence peut n'avoir aucune valeur diagnostique ou même être associée à des effets inhibiteurs sur d'autres régions des objets discutés ou sur d'autres objets.

### **Attributs de niveau intermédiaire : géons, squelette et surfaces**

Les attributs de niveau intermédiaire ne ressortent pas de manière évidente dans les IC moyennes. Il n'y a qu'un seul cas qui semble définir une telle composante (soit un géon ou un segment du squelette de l'objet), soit l'objet 2. Il pourrait s'agir soit d'un géon (Biederman, 1987) ou du squelette de l'objet (Feldman et al., 2013; Leyton 1989) provenant de l'information des contours et sur la base duquel les axes médians seront ensuite inférés.

Dans tous les autres cas, ce qu'on pourrait définir comme une composante de l'objet est occupé à la fois par des régions ayant un effet facilitateur et d'autres ayant un effet inhibiteur. Ces juxtapositions semblent incompatibles avec une représentation des objets en termes de composantes (Marr & Nishihara, 1978; Marr, 1982; Biederman, 1987; Pentland, 1986; Feldman & al, 2013; Leyton, 1989). L'idée d'une représentation basée sur les surfaces (Leek, Reppa & Arguin 2005) souffre autant à l'analyse. Il y a très peu de régions formant une surface continue qui n'est pas interrompue par une région inhibitrice parmi celles ayant un effet facilitateur pour la reconnaissance.

## **Fragments d'image**

Tel qu'illustré à la Fig. 4, les régions efficaces que démontrent les IC sont compatibles avec ce que propose la théorie des fragments qui, de fait ne pose aucune contrainte particulière quant aux propriétés caractérisant les régions de l'image susceptible d'être efficaces pour reconnaître un objet, hormis leur valeur informative. On y observe une variété d'attributs élémentaires de divers types (voir section précédente à cet effet), ayant une haute valeur diagnostique pour l'identification de certaines cibles. Donc, il semble possible que dans le cas où ils ne disposeraient que des régions illustrées en rouge/orange sur les IC moyennes, les participants pourraient néanmoins identifier les objets de façon fiable, i.e. avec un niveau de succès très élevé. Un aspect se dégageant du contenu des zones efficaces, et ce peu importe l'objet, est que leur localisation semble n'avoir que peu d'importance. Ainsi, il est apparent que les zones efficaces qui sont hautement diagnostiques pour la reconnaissance des objets leurs sont uniques; autrement dit, elles ne sont pas partagées avec d'autres objets. Suivant ce raisonnement, nous pourrions varier aléatoirement la localisation des stimuli sur l'écran d'un essai à l'autre plutôt que de les centrer sur le point de fixation comme nous avons fait, et néanmoins répliquer les images de classification observées ici.

## Conclusion

Dans le cadre de cette étude, nous avons identifié les zones importantes de l'image (i.e. celles ayant un effet facilitateur ou inhibiteur) pour la reconnaissance d'un ensemble de 8 objets 3D non familiers faits de deux composantes par un groupe de 12 observateurs humains normaux. Les résultats obtenus indiquent que des attributs élémentaires hétérogènes caractérisent les régions hautement diagnostiques de l'image. Aucun des types d'attributs élémentaires considérés ne paraît posséder un pouvoir diagnostique intrinsèquement plus important que les autres. La sélection des attributs reposerait plutôt sur leur valeur informative pour l'identification correcte de la cible. Pour être sélectionnés, les attributs doivent être en mesure d'être encodés et parmi eux, ceux qui paraissent à la fois simples et distinctifs ont été utilisés pour discriminer entre les objets de l'ensemble. Advenant qu'une stratégie élémentaire s'avère inefficace par exemple, avec un ensemble d'objets plus vaste ou plus similaires entre eux, il est probable qu'une stratégie plus complexe puisse être appliquée. À travers l'ensemble des théories de la reconnaissance d'objets envisagé dans l'Introduction, notre analyse des résultats obtenus semble avoir la plus grande compatibilité avec la théorie de reconnaissance basée sur les fragments informatifs.

*Fin*

---



## Références bibliographiques

- Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*(1), 703-711. doi:<https://doi.org/10.1016/j.neuroimage.2009.08.046>
- Arguin, M., & Saumier, D. (2004). Independent processing of parts and of their spatial organization in complex visual objects *15*(9), 629-633. *Psychological Science*, *15*(9), 629-633. doi:<https://doi.org/10.1111/j.0956-7976.2004.00731.x>
- Arguin, M., Ferrandez, R., & Massé, J. (2021). Oscillatory Visual Mechanisms Revealed by Random Temporal Sampling. *ResearchSquare*. doi:<https://doi.org/10.21203/rs.3.rs-244360/v1>
- Ban, H., Yamamoto, H., Hanakawa, T., Urayama, S., Aso, T., Fukuyama, H., & Ejima, Y. (2013). Topographic representation of an occluded object and the effects of spatiotemporal context in human early visual areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(43), 16992–17007. doi:<https://doi.org/10.1523/JNEUROSCI.1455-12.2013>
- Barenholtz, & Tarr. (2011). Visual Learning of Statistical Relations Among Non-adjacent Features: Evidence for Structural Encoding. *Visual cognition*, *19*(4), 469–482. doi:<https://doi.org/10.1080/13506285.2011.552894>
- Basri, R., & Ullman, S. (1991). Linear operator for object recognition. *Advances in Neural Information Processing Systems*, *4*, 452-459.
- Bertramini, M., & Wagemans, J. (2013). Processing convexity and concavity along a 2-D contour: figure-ground, structural shape, and attention. *20*, 191-207. doi:<https://doi.org/10.3758/s13423-012-0347-2>
- Biederman, I. (1987). Recognition by components: a theory of visual pattern recognition. *Psychological Review*, *94*, 115-147. doi:[https://doi.org/10.1016/S0079-7421\(08\)60015-8](https://doi.org/10.1016/S0079-7421(08)60015-8)
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1162-1182. doi:10.1037//0096-1523.19.6.1162
- Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, *20*(1), 38-64. doi:[https://doi.org/10.1016/0010-0285\(88\)90024-2](https://doi.org/10.1016/0010-0285(88)90024-2)

- Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838. doi:<https://doi.org/10.1016/j.neuropsychologia.2012.08.010>.
- Breakspear, M., Brammer, M. J., Bullmore, E. T., Das, P., & Williams, L. M. (2004). Spatiotemporal wavelet resampling for functional neuroimaging data. *Human brain mapping*, *23*(1), 1-25. doi:<https://doi.org/10.1002/hbm.20045>
- Breakspear, M., Brammer, M., & Robinson, P. A. (2003). Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. *Physica D: Nonlinear Phenomena*, *182*(1-2), 1-22. doi:[https://doi.org/10.1016/S0167-2789\(03\)00136-2](https://doi.org/10.1016/S0167-2789(03)00136-2).
- Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*(3), 247-260. doi:<https://doi.org/10.1093/cercor/5.3.247>
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 6.59-667. doi:<https://doi.org/10.1167/5.9.1>
- Coggan, D. D., Baker, D. H., & Andrews, T. J. (2019). Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal of Neuroscience*, *49*(12), 1587-1596. doi:<https://doi.org/10.1111/ejn.14327>
- Cortese, J. M., & Dyre, B. P. (1996). Perceptual similarity of shapes generated from fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 133. doi:10.1037/0096-1523.22.1.133
- Crouzet, S. M., & Serre, T. (2011). What are the visual features underlying rapid object recognition? *Frontiers in psychology*, *2* art. no. 326. doi:<https://doi.org/10.3389/fpsyg.2011.00326>
- Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering*, *36*(1), 107-114. doi:10.1109/10.16456
- Duda, R., & al. (2000). *Pattern Classification* (éd. 2). John Wiley & Sons.
- Edelman, S., & Weinshall, D. 6. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, 209-219.
- Edelman, S., Bulthoff, H., & Weinshall, D. (1989). *Stimulus familiarity determines recognition strategy for novel 3D objects (Artificial Intelligence Laboratory Technical Report No. 1138)*. Cambridge, MA: MIT Press.
- Edelman, S., & Bulthoff, H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision research*, *32*, 2385-2400. doi:[https://doi.org/10.1016/0042-6989\(92\)90102-o](https://doi.org/10.1016/0042-6989(92)90102-o)

- Epshtein, B., & Ullman, S. (2005). Feature hierarchies for object classification., *1*, pp. 220–227.
- Epshtein, B., & Ullman, S. (2006). Satellite features for the classification of visually similar classes., *2*, pp. 2079–2086. doi:10.1109/CVPR.2006.262
- Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. Dans I. S. (Eds), *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective* (pp. 55-70). doi:https://doi.org/10.1007/978-1-4471-5195-1\_4
- Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*, 771-781. doi:https://doi.org/10.1016/j.visres.2010.09.027.
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. doi:https://doi.org/10.1016/S0042-6989(01)00097-9
- Harel, A., Ullman, S., Epshtein, B., & Bentin, S. (2007). Mutual information of image fragments predicts categorization. *Vision research*, *47*(15), 2010-2020. doi:https://doi.org/10.1016/j.visres.2007.04.004
- Harel, A., Ullman, S., Harari, D., & Bentin, S. (2011). Basic-level categorization of intermediate complexity fragments reveals top-down effects of expertise in visual perception. *Journal of vision*, *11*(8), 18. doi:https://doi.org/10.1167/11.8.18
- Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *15*11-1521. doi:https://doi.org/10.1037/0096-1523.23.5.1511
- Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*, *11*(1), 7-12. doi:https://doi.org/10.1111/1467-9280.00207
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive Fields and Functional Architecture of Monkey Striate Cortex. *Journal of Physiology*. *Journal of Physiology*, 215-243. doi:10.1113/jphysiol.1968.sp008455
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1-16.
- Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., Von Der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture . *IEEE Transactions on computers*, *42*(3), 300-311. doi:https://doi.org/10.1109/
- Leek, E. C., Reppa, I., & Arguin, M. (2005). The structure of 3D object shape representations: Evidence from whole-part matching. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 668-684. doi:https://doi.org/10.1080/17470210802303826

- Leek, E. C., Reppa, I., Rodriguez, E., & Arguin, M. (2009). Surface but not volumetric part structure mediates three dimensional shape representation: Evidence from part-whole priming. *Quarterly Journal of Experimental Psychology*, 62(4), 814-830. doi:<https://doi.org/10.1080/17470210802303826>
- Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, 13(3), 357-387. doi:[https://doi.org/10.1016/0364-0213\(89\)90017-7](https://doi.org/10.1016/0364-0213(89)90017-7)
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, CA, U.S.A: Freeman., W.H. doi:10 9 8 7 6 5 4 3 2 1
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological sciences*, 200 (140), pp. 269–294. London. doi:  
<http://www.jstor.org/stable/77391>
- McCotter, Gosselin, Sowden, & Schyns. (2005). The use of visual information in natural scenes. *Visual Cognition*, 12(6), 938-953. doi:<https://doi.org/10.1080/13506280444000599>
- Pasupathy, A., & Connor, C. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5,, 5(12), 1332-1338. doi:<https://doi.org/10.1038/nn972>
- Pentland., A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28(3), 293-331. doi:[https://doi.org/10.1016/0004-3702\(86\)90052-4](https://doi.org/10.1016/0004-3702(86)90052-4).
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343, 269-266. doi:<https://doi.org/10.1038/343263a0>
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11), 1019-1025. doi:<https://doi.org/10.1038/14819>
- 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. doi:10.1111/1467-9280.00472
- Schyns, P., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, 5(4), 195-200. doi:<https://doi.org/10.1111/j.1467-9280.1994.tb00500.x>
- Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science*, 12(6), 437-444. doi:<https://doi.org/10.1111/1467-9280.00382>
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences USA*, 104, pp. 6424-6429.

- Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *New Series*, 171(3972), 701-703. Récupéré sur <https://www.jstor.org/stable/1731476>
- Smith, A., Singh, K., Williams, A., & Greenlee, M. (2001). Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cerebral Cortex*, 11(12), 1182-1190. doi: <https://doi.org/10.1093/cercor/11.1>
- Stankiewicz, B. J., & Hummel, J. E. (1996). Categorical relations in shape perception. *Spatial vision*, 10(3), 201-236.
- Tarr, M. ..., Williams, P., Hayward, G. W., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nat Neurosci*, 1, 275-277. doi: <https://doi.org/10.1038/1089>
- Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, 21(6), 1494–1505. doi:<https://doi.org/10.103>
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive psychology*, 21(2), 233-282. doi:[https://doi.org/10.1016/0010-0285\(89\)90009-1](https://doi.org/10.1016/0010-0285(89)90009-1)
- Ullman. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 193-224. doi:[10.1016/0010-0277\(89\)90036-X](https://doi.org/10.1016/0010-0277(89)90036-X)
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in cognitive sciences*, 11(2), 58-64. doi:<https://doi.org/10.1016/j.tics.2006.11.009>
- Ullman, S., & Bart, E. (2004). Recognition invariance obtained by extended and invariant features. *Neural Networks*, 17(5-6), 833-848. doi:<https://doi.org/10.1016/j.neunet.2004.01.006>
- Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences of the United States of America*, 2744-2749. doi:<https://doi.org/10.1073/pnas.1513198113>
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature neuroscience*, 682-687. doi:<https://doi.org/10.1038/nn870>
- Uyar, F., Shomstein, S., Greenberg, A. S., & Behrmann, M. 9.-1. (2016). Retinotopic information interacts with category selectivity in human ventral cortex. *Neuropsychologia*, 92, 90-106. doi:<https://doi.org/10.1016/j.neuropsychologia.2016.05.022>
- Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3D object recognition: invariance to imaging transformations. *Cerebral Cortex*, 5(3), 261-269. doi:<https://doi.org/10.1093/cercor/5.3.261>

- Wiesmann, S. L., Caplette, L., Willenbockel, V., Gosselin, F., & Võ, M. L. (2021). Flexible time course of spatial frequency use during scene categorization. *Scientific Reports, 11*(1), 1-13.
- Willenbockel, V., Sadr, J., & Fiset, D. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*(3), 671-684.  
doi:<https://doi.org/10.3758/BRM.42.3.671>