



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Marlène Poncet

Le vendredi 6 juin 2014

Titre :

Mécanismes cognitifs dans la catégorisation d'objets visuels

École doctorale et discipline ou spécialité :

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

CerCo - Centre de Recherche Cerveau et Cognition

Directeur(s) de Thèse :

Michèle Fabre-Thorpe (CerCo, Toulouse)

Rapporteurs :

Aude Oliva (MIT, Cambridge, États-Unis)

Olivier Pascalis (LPNC, Grenoble)

Autre(s) membre(s) du jury :

Jonathan Grainger (LPC, Marseille)

Guillaume Rousselet (University of Glasgow, Royaume-Uni)

Martin Giurfa (CRCA, Toulouse)



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Marlène Poncet

Le vendredi 6 juin 2014

Titre :

Mécanismes cognitifs dans la catégorisation d'objets visuels

École doctorale et discipline ou spécialité :

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

CerCo - Centre de Recherche Cerveau et Cognition

Directeur(s) de Thèse :

Michèle Fabre-Thorpe (CerCo, Toulouse)

Rapporteurs :

Aude Oliva (MIT, Cambridge, États-Unis)

Olivier Pascalis (LPNC, Grenoble)

Autre(s) membre(s) du jury :

Jonathan Grainger (LPC, Marseille)

Guillaume Rousselet (University of Glasgow, Royaume-Uni)

Martin Giurfa (CRCA, Toulouse)

Table des matières

1. Les représentations des catégories d'objets.....	7
1.1. Où ? La voie visuelle ventrale : siège de la reconnaissance des objets	9
1.1.1. Activités évoquées par des stimuli particuliers	11
1.1.2. Activité sélective et rôle fonctionnel.....	15
1.1.3. Vers une représentation distribuée des catégories d'objets.....	18
1.1.4. Nouvelles idées, nouvelles techniques, nouveaux résultats	20
1.1.5. Résumé	30
1.2. Quand ? Les catégories d'objets visuels sont accessibles en une centaine de millisecondes.....	31
1.2.1. Temps de présentation minimum pour avoir accès aux informations d'une scène visuelle	31
1.2.2. Temps de réponse minimum pour déterminer la catégorie d'un objet.....	34
1.2.3. Temps à partir duquel l'activité neuronale permet de différencier différentes catégories d'objets visuels	37
1.2.4. Résumé	41
1.3. Comment ? Modèles du traitement de l'information visuelle.....	43
1.3.1. Organisation hiérarchique des catégories d'objets	43
1.3.2. Une première vague feed-forward pour expliquer la catégorisation rapide	46
1.3.3. Résumé	50
1.4. Résumé de l'introduction.....	51
2. Rapidité d'accès aux différents niveaux de catégories	53
2.1. Un objet, plusieurs catégories	54
2.1.1. La catégorie basique : point d'entrée de la reconnaissance visuelle	55
2.1.2. La catégorisation basique n'est pas une étape nécessaire à la reconnaissance visuelle d'objets	57
2.1.3. Présentation de l'article 1	59
2.2. Article 1: Stimulus duration and diversity do not reverse the advantage for superordinate representations: the animal is seen before the bird	60
2.3. Discussion	72
2.3.1. Quel niveau de représentation est accédé en premier ?.....	73
2.3.2. Importance de la discriminabilité dans la catégorisation	74
2.3.3. Effet bloc ?.....	76

3. Rôle de l'attention dans la catégorisation d'objets visuels	79
3.1. L'étude de l'attention	80
3.1.1. Des stimuli simples.....	81
3.1.2. ... aux stimuli complexes	83
3.1.3. La théorie de l'intégration des traits revisitée.....	85
3.1.4. Présentation de l'article 2.....	91
3.2. Article 2: A need for more information uptake but not focused attention to access basic-level representations.....	93
3.3. Discussion	111
3.3.1. Quel type de représentation est créé sans attention ?.....	111
3.3.2. Que veut dire en quasi-absence d'attention ?.....	112
4. Encodage à long terme pendant la catégorisation d'objets	115
4.1. Niveau de traitement des images lors de catégorisations visuelles rapides	116
4.1.1. Entraînement sur le même set d'images familières	118
4.1.2. Présentation de l'article 3	119
4.2. Article 3: Briefly flashed scenes can be stored in long-term memory ...	121
4.3. Discussion	149
5. Interactions entre les catégories visuelles	153
5.1. Présentation de la problématique	155
5.1.1. Etudier les interactions entre des représentations.....	155
5.1.2. ...en relation avec l'organisation des catégories dans IT	157
5.1.3. Présentation de l'article 4.....	158
5.2. Article 4: The clash of visual categories	159
5.3. Discussion	186
6. Discussion générale	190
6.1. Résumé des résultats.....	191
6.1.1. Etude 1.....	191
6.1.2. Etude 2.....	192
6.1.3. Etude 3.....	193
6.1.4. Etude 4.....	195
6.2. Implications	198
6.2.1. D'où vient le décalage temporel ?	200

6.2.1.	Comment sont codées les caractéristiques ?.....	204
6.2.2.	Un rôle pour l'attention dans la catégorisation d'objets?	206
6.3.	Limitation : les caractéristiques des objets... de quoi parle-t-on exactement?.....	208
6.4.	Perspectives.....	211
6.4.1.	Quelle est la localisation cérébrale de la compétition entre représentations visuelles ?	211
6.4.2.	Quelle est la nature des caractéristiques conduisant à une compétition entre les représentations visuelles ?	212
6.4.3.	Représentations visuelles ou/et représentations conceptuelles ?	212
7.	Bibliographie	213
8.	Annexes.....	234
8.1.	Liste des acronymes	234
8.2.	Poster Mack & Palmeri (2011b).....	235
8.3.	Analyses of “no” responses in study 4.....	237
8.3.1.	Experiment 1	237
8.3.2.	Experiment 2	239
8.3.1.	Summary	241



Chapitre 1
Introduction



1. Les représentations des catégories d'objets

Comment parvenons-nous à reconnaître si rapidement et sans effort l'environnement complexe qui nous entoure ? Cet acte apparemment si anodin, à la base de nos activités quotidiennes, est pourtant très complexe. En un clin d'œil nous pouvons reconnaître une scène visuelle en 3D et en mouvement. Nous reconnaissons des objets de formes, tailles et couleurs variées même si ils sont en partie cachés. Nous percevons des odeurs, des sons, des sensations, etc. Comment le cerveau réussit-il à former une représentation cohérente de toutes ces informations en si peu de temps ? La vision est le sens utilisé majoritairement pour interagir avec notre environnement. Environ 20 à 25% du cortex humain est dédié au traitement de l'information visuelle. Même si les autres modalités sensorielles sont aussi importantes et interagissent avec la vision (e.g. effet Mc Gurk), j'ai choisi d'étudier dans ma thèse comment le cerveau traite les informations visuelles, laissant de côté les interactions possibles avec les autres sens.

Chaque fois que nous voyons un objet, cet objet est unique : il n'a peut-être pas la même orientation, n'est pas éclairé de la même manière, n'a pas la même couleur, est caché par un autre objet, etc. Et pourtant, nous avons besoin de toujours l'identifier comme étant le même objet. Le système visuel doit donc établir une équivalence des différents patterns de réponses de la rétine et dans le même temps ne pas confondre cet objet avec un autre. La capacité à rattacher à un même objet les diverses perceptions qu'on peut en avoir, pouvoir le distinguer ou le regrouper avec différents objets, est une des aptitudes de l'humain (et d'autres animaux, e.g. Fabre-Thorpe, Richard, & Thorpe, 1998; Herrnstein & Loveland, 1964; Vogels, 1999a) la plus importante pour interagir avec son environnement. Le monde étant riche d'une grande variabilité, il nous est nécessaire de regrouper un grand nombre d'images sous le même label pour ne pas surcharger notre mémoire tout en étant capables de nous adapter très facilement à un nouvel environnement composé d'objets jamais rencontrés auparavant. La catégorisation, c'est-à-dire la capacité à encoder des caractéristiques générales des objets afin de pouvoir les classer au-delà de leurs propriétés sensorielles spécifiques, est une des questions majeures en neurosciences et ses mécanismes restent largement mystérieux.

Au cours de cette thèse, je me suis principalement intéressée aux mécanismes cognitifs impliqués dans la catégorisation d'objets en essayant de comprendre dans quatre études (1) quels sont les paramètres

influençant les performances de catégorisation, (2) le rôle de l'attention dans la catégorisation d'objets, (3) le lien entre la catégorisation rapide et la mémorisation, et (4) les interactions entre les catégories d'objets.

Je présente dans cette première partie les grandes lignes des connaissances accumulées dans le domaine de la reconnaissance visuelle et plus particulièrement concernant les catégories d'objets. J'ai séparé cette introduction en trois chapitres, chacun traitant de trois questions différentes sur le traitement des catégories d'objets : où a-t-il lieu (chapitre 1.1), en combien de temps (chapitre 1.2), comment (chapitre 1.3) ?

1.1. Où ? La voie visuelle ventrale : siège de la reconnaissance des objets

Des années de recherches en neurosciences ont permis de fournir de nombreux résultats confirmant le rôle très important de la voie visuelle ventrale dans l'identification des objets visuels (voir revue de Gross, 1994). Les premiers travaux fondamentaux ont été ceux de Ungerleider & Mishkin, (1982) qui ont proposé deux voies différentes du traitement de l'information visuelle. Ils ont décrit une voie visuelle ventrale dédiée au traitement des objets distincte d'une voie visuelle dorsale, dédiée au traitement de la vision spatiale (Goodale & Milner, 1992; Ungerleider & Haxby, 1994).

Felleman & Van Essen, (1991) ont proposé une organisation hiérarchique des aires visuelles, opposée à une organisation en parallèle. Ils ont déduit par l'observation des patterns de connectivité, que l'information visuelle captée par la rétine et relayée par les corps genouillés latéraux du thalamus (CGL)¹, était transmise successivement dans les aires corticales de la voie ventrale de V1 à V2 à V4 puis vers le cortex inféro-temporal (IT). Ces résultats ont été confirmés par la suite dans des revues montrant une augmentation des latences des réponses d'environ 10 ms étape après étape entre V1 et IT (Nowak & Bullier, 1997; Schmolesky et al., 1998). Ces résultats sont résumés et illustrés dans la Figure 1.

Le traitement de l'information visuelle dans les premières étapes de la voie visuelle ventrale est assez bien connu. Les neurones de V1 présentent une sélectivité pour des attributs simples qui couvrent tous les aspects de la vision : forme, mouvement, couleur, texture et profondeur (Lennie & Movshon, 2005; Orban, 2008). Les étapes les plus tardives (V4 et IT) sont moins bien connues.

¹ La liste des acronymes est fournie en annexe.

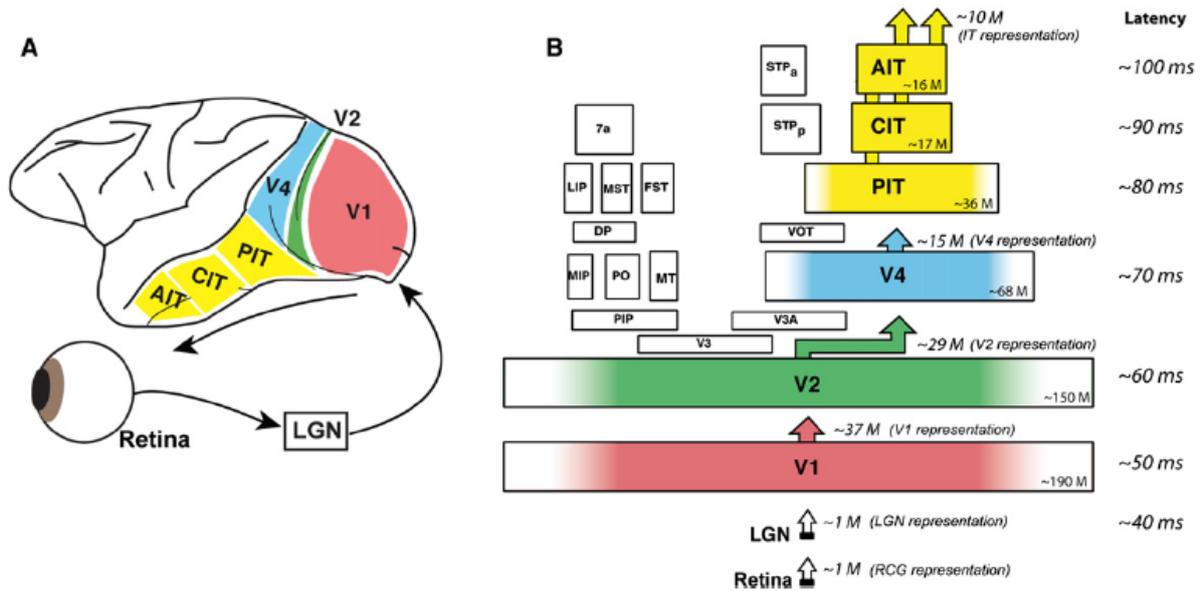


Figure 1. (A) Flux de l'information visuelle de la rétine jusque dans les aires de la voie ventrale chez le macaque. (B) Chaque aire est représentée de manière à ce que sa taille soit proportionnelle à sa surface corticale (Felleman & Van Essen, 1991). Le nombre total de neurones (dans les 2 hémisphères) est donné au-dessus de chaque aire. Les latences médianes sont listées sur la droite (Nowak & Bullier, 1997; Schmolesky et al., 1998). (DiCarlo, Zoccolan, & Rust, 2012)

Des lésions dans les aires postérieures de la voie visuelle ventrale engendrent une cécité dans certaines régions du champ visuel. Au contraire, des lésions ou l'inactivation des régions antérieures de la voie visuelle ventrale, en particulier dans IT produisent des déficits sélectifs dans la capacité à discriminer des objets complexes. Par exemple chez le singe, des lésions du cortex extrastrié, de V4 ou de IT entraînent des déficits dans la reconnaissance d'un objet présenté dans une vue non prototypique (objet ayant subi des transformations de taille, de degré d'occlusion, etc.), alors que l'identification du même objet dans une vue prototypique est préservée (Holmes & Gross, 1984; Schiller, 1995; Elizabeth K. Warrington & Taylor, 1978; Weiskrantz & Saunders, 1984; Yaginuma, Niihara, & Iwai, 1982). Chez l'homme, des patients atteints au niveau de la voie visuelle ventrale ont des difficultés à reconnaître des objets visuels (agnosie visuelle, Biran & Coslett, 2003). Toutefois, ils peuvent reconnaître le même objet par d'autres sens tels que le toucher, l'audition ou l'odorat. Ils sont capables de le dessiner ou de l'utiliser correctement.

Les déficits de reconnaissance des objets visuels peuvent parfois toucher des catégories d'objets spécifiques. Un des exemples le plus connu est celui de la prosopagnosie : les patients sont incapables de reconnaître des visages familiers tout en restant capables de reconnaître d'autres objets visuels (Benton, 1980; Damasio, Damasio, & Van Hoesen, 1982; De Renzi, 1986; Hecaen & Angelergues, 1962; McNeil & Warrington, 1993; Sergent & Signoret, 1992). D'autres études ont rapporté le cas de patients ayant des performances déficitaires dans la reconnaissance d'objets vivants, mais préservées dans le cas d'objets non vivants (Farah, McMullen, & Meyer, 1991; Elizabeth K. Warrington & McCarthy, 1983; Elizabeth K. Warrington & Shallice, 1984). Ces types de déficit observés chez l'humain ont fait émerger l'hypothèse que la reconnaissance d'objets pourrait s'appuyer sur des mécanismes spécifiques au traitement de catégories d'objets particulières.

IT serait donc le premier lieu où l'information de la catégorie d'un objet serait implémentée. La thématique de la thèse portant sur l'accès aux catégories des objets, on s'intéressera plus précisément aux informations contenues dans IT. On débutera cette introduction avec la découverte de neurones ou d'aires corticales répondant préférentiellement à un type de stimulus particulier (chapitre 1.1.1) pour s'interroger ensuite sur leur fonction (chapitre 1.1.2). Dans les deux derniers chapitres, on développera l'idée d'une organisation distribuée des représentations (chapitre 1.1.3) ainsi que les résultats mis en évidence avec les nouvelles techniques d'analyses (analyse multi-variée de patterns, MVPA) qui prennent en compte cette organisation distribuée (chapitre 1.1.4).

1.1.1. Activités évoquées par des stimuli particuliers

1.1.1.1 Enregistrements de réponses neuronales invariantes

Konorski, (1967) est probablement la première personne à avoir émis l'hypothèse de l'existence de cellules « gnostiques » dans le cerveau qui représenteraient des perceptions unitaires. Il a même suggéré que de telles cellules pourraient être présentes dans IT et coder pour les visages, les mains et les expressions faciales. Des neurones de ce type ont été reporté pour la première fois par Gross, Bender, & Rocha-Miranda, (1969). Les neurones qu'ils avaient observés déchargeaient beaucoup plus lors de la présentation d'une main que lors de la présentation d'autres formes. Trois ans plus tard, les auteurs

ont retrouvé les mêmes neurones sélectifs à la forme d'une main et ont aussi enregistré pour la première fois des neurones sélectifs aux visages (Gross, Rocha-Miranda, & Bender, 1972).

Ces découvertes ont été marquées de beaucoup de scepticisme. Il faudra attendre une dizaine d'années pour que d'autres chercheurs enregistrent eux aussi des populations de neurones très fortement activées pour les visages. Les réponses de ces neurones dépendaient de la configuration spécifiques des traits d'un visage et leur sélectivité était maintenue pour des stimuli de taille, de position ou d'éclaircissement différentes (Bruce, Desimone, & Gross, 1981; Desimone, Albright, Gross, & Bruce, 1984; Perrett, Rolls, & Caan, 1982).

Des travaux similaires utilisant différents types de stimuli (des formes géométriques jusqu'à des objets réels) ont pu montrer que la plupart des neurones de IT ont une activité préférée pour un objet particulier. Leur sélectivité était maintenue quelles que soient la taille, la position (Brincat & Connor, 2004; Ito, Tamura, Fujita, & Tanaka, 1995; Tovée, Rolls, & Azzopardi, 1994), l'orientation (Logothetis, Pauls, Bülthoff, & Poggio, 1994), l'éclaircissement (Vogels & Biederman, 2002) de l'objet, ou si d'autres objets étaient présentés en même temps (Missal, Vogels, Li, & Orban, 1999; Missal, Vogels, & Orban, 1997; Zoccolan, Cox, & DiCarlo, 2005).

Ces études montrent l'existence de neurones de IT sélectifs et plus important encore, *invariants* aux modifications de l'objet. Elles ont été conduites chez le singe mais il est probable que de tels neurones puissent aussi être présents chez l'humain. En effet, les propriétés et l'organisation de IT semble très bien correspondre au cortex occipital latéral (LOC) chez l'humain (Kriegeskorte et al., 2008; Orban, Van Essen, & Vanduffel, 2004).

1.1.1.2 Régions sélectives (IRMf)

Avec l'imagerie par résonance magnétique fonctionnelle (IRMf, Kwong et al., 1992; Ogawa et al., 1992), de nombreux travaux ont cherché l'existence d'aires cérébrales dédiées au traitement d'objets spécifiques chez l'humain. Une de ces premières études a montré que LOC était préférentiellement activé par des images d'objets quels qu'ils soient, par rapport à des textures (Malach et al., 1995). Plus intéressant, les régions du cortex situées autour de LOC se sont révélées sélectives à des catégories d'objets particulières : elles répondaient au moins deux fois plus pour leur stimulus 'préféré' que pour

d'autres stimuli. Il est intéressant de noter que la localisation et les propriétés fonctionnelles de ces régions sont très similaires entre les participants (Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher & Yovel, 2006).

Ainsi, différentes régions répondant sélectivement à différents types de stimuli ont été trouvées (Figure 2). Parmi elles, des régions sélectives aux visages (dans le gyrus fusiforme des visages « FFA », Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996 ; et le cortex occipital inférieur des visages, Rossion et al., 2003) et d'autres aux parties du corps (dans l'aire extrastrée des corps « EBA », Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Downing, Jiang, Shuman, & Kanwisher, 2001; Peelen & Downing, 2007 ; et le gyrus fusiforme des corps, Schwarzlose, Baker, & Kanwisher, 2005; Taylor, Wiggett, & Downing, 2007). En ce qui concerne le traitement des informations des scènes visuelles, une partie de l'aire parahippocampique, serait dédiée au traitement des lieux et des bâtiments (Aguirre, Zarahn, & D'esposito, 1998; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). Une autre région du cortex inférieur occipitotemporal (aire de la forme des mots visuels) a aussi été identifiée pour le traitement des mots (Baker et al., 2007; Ben-Shachar, Dougherty, Deutsch, & Wandell, 2007; L. Cohen et al., 2000).

L'étude de l'activité corticale dans son ensemble montre que les régions médianes de la surface de la voie visuelle ventrale (le gyrus fusiforme médian, le gyrus lingual, et le cortex parahippocampique) ont une activité plus importante pour les artefacts comme les outils et les objets non manipulables par rapport aux objets animés tels que les animaux et les visages. Au contraire, les régions latérales de la surface de la voie visuelle ventrale (le gyrus fusiforme latéral, le gyrus temporal) ont une activité plus importante pour les objets vivants par rapport aux artefacts (en IRMf : Chao, Haxby, & Martin, 1999; Downing et al., 2006 ; et en tomographie par émission de positons, TEP, Martin, Wiggs, Ungerleider, & Haxby, 1996; Perani et al., 1995).

Des régions sélectives pour toutes les catégories d'objets n'ont pas été observées, ce qui ne permet cependant pas de conclure qu'elles n'existent pas. En effet, il est possible que la différence d'activité entre deux catégories de stimuli ne soit pas visible avec la résolution spatiale actuelle de l'IRMf mais qu'avec une résolution plus fine, nous puissions distinguer ces différences d'activité.

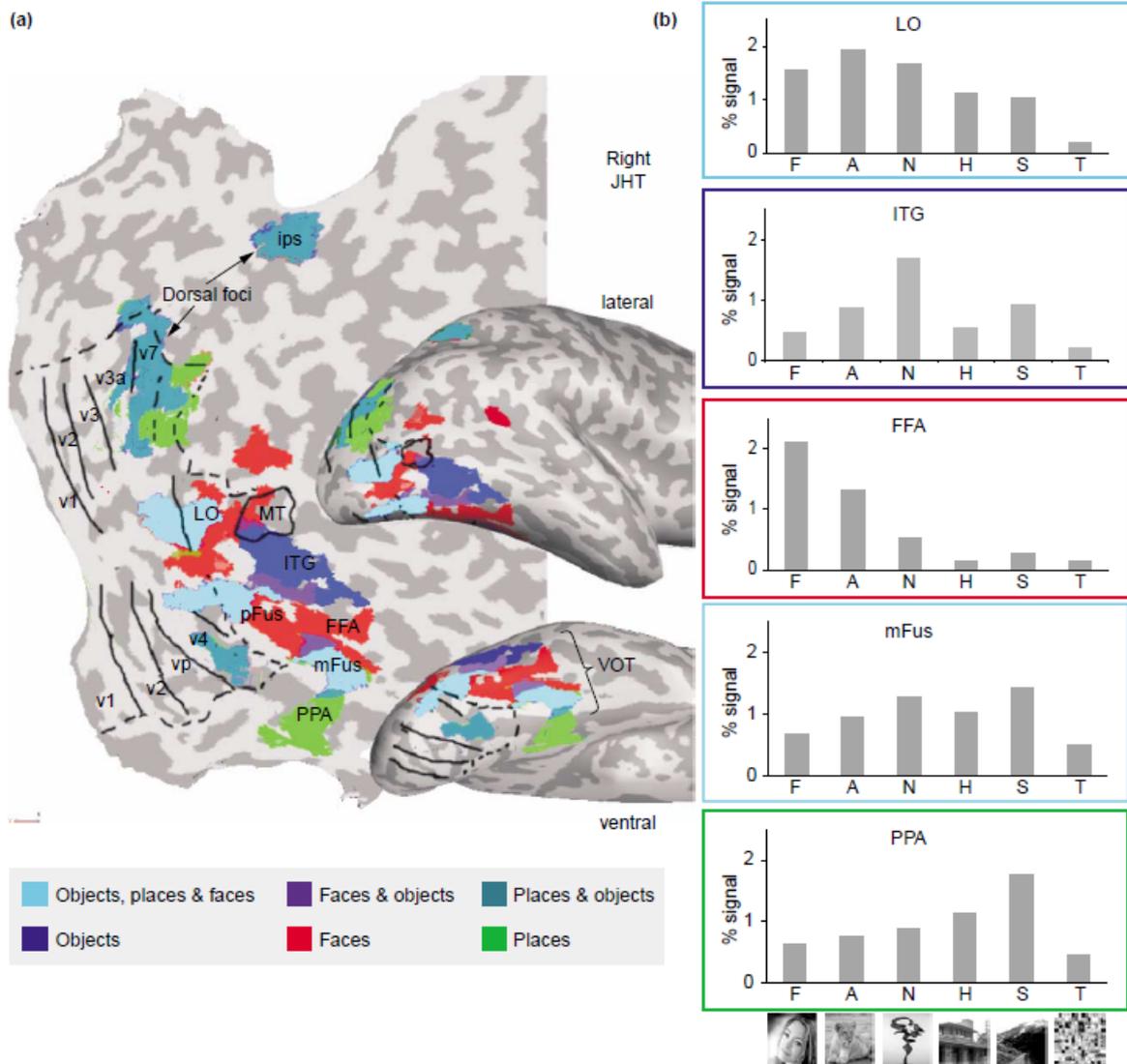


Figure 2. Régions sélectives à certains objets visuels dans le cerveau humain. (a) Régions activées plus fortement pour des images d'objets (chaque type d'objet correspond à une couleur différente) que pour ces mêmes images d'objets mais brouillées (des exemples des stimuli utilisés sont représentés en bas à droite de la figure 2, l'initiale de chaque objet est utilisée dans les résultats 2b). (b) Pourcentage de changement de signal pour différentes régions d'intérêt (LO: aire occipito-latérale; ITG : gyrus inféro-temporal, FFA : aire fusiforme des visages, mFus : mi-fusiforme, PPA : aire parahippocampique des lieux). (Grill-Spector, 2003)

1.1.2. Activité sélective et rôle fonctionnel

L'activité préférentielle d'un neurone ou d'une région cérébrale pour un stimulus particulier ne veut pas obligatoirement dire que ce neurone ou cette région joue un rôle dans la perception de ce stimulus. La corrélation entre perception et augmentation préférentielle d'activité a été mise en évidence en utilisant différentes méthodes dont les résultats sont développés ensuite : microstimulation de populations de neurones induisant un biais comportemental (chapitre 1.1.2.1), corrélation entre performance de reconnaissance et niveau d'activité de certaines régions cérébrales (chapitre 1.1.2.2), perturbation de régions cérébrales en utilisant la stimulation magnétique transcrânienne (TMS) (chapitre 1.1.2.3).

1.1.2.1 Rôle de l'activité préférentielle des neurones de IT dans la perception

Récemment, des travaux ont pu montrer que l'activation artificielle des neurones de IT chez le singe induisait des biais dans la perception d'objets complexes (Afraz, Kiani, & Esteky, 2006). Les chercheurs ont activé artificiellement (par micro-stimulations électriques) des groupes de neurones dans IT pendant que les singes réalisaient une tâche de catégorisation visage vs. non-visages en utilisant des images bruitées. La micro-stimulation des régions sélectives aux visages biaisait fortement la décision des singes en faveur de la catégorie visage, ce qui n'était pas le cas lors de la stimulation d'autres régions corticales. L'amplitude de cet effet augmentait avec (a) l'augmentation du degré de sélectivité pour les visages des neurones et (b) l'augmentation de la taille du cluster de neurones. Le biais comportemental provoqué par la micro-stimulation était maximal lorsque la stimulation était appliquée vers 50-100 ms après le début de la présentation de l'image, ce qui suggère que ces informations très précoces sont utilisées pour la prise de décision. La micro-stimulation avait un effet moindre lorsqu'elle était appliquée entre 100-150 ms et n'avait pas d'effet significatif lorsqu'elle était appliquée entre 0-50 ms. Ces résultats établissent une relation causale entre l'activité des neurones sélectifs aux visages et la perception des visages.

1.1.2.2 Perception consciente et niveau d'activité de régions corticales

En utilisant un paradigme de masquage en IRMf, Grill-Spector, Kushnir, Hendler, & Malach, (2000) ont montré que les performances d'identification des participants étaient corrélées avec le niveau d'activation des aires sélectives aux objets dans le cortex visuel ventral. Les auteurs présentaient des images d'objets que des participants devaient nommer. Le temps de présentation des stimuli variait entre 20 et 500ms avant l'apparition d'un masque. Les participants identifiaient correctement presque toutes les images lorsqu'elles étaient présentées pendant 500 ms mais leurs performances diminuaient avec la diminution du temps de présentation de l'objet jusqu'à ce que les participants soient incapables d'identifier l'objet pour des présentations de 20 ms. Les essais dans lesquels les participants arrivaient à nommer le stimulus produisaient une activation cérébrale plus forte (en particulier dans les aires latérales occipitales) que lorsque l'objet était présenté mais non identifié par le participant. De plus, en utilisant le même type de paradigme mais en permettant l'amélioration des performances d'identification grâce à la répétition des stimuli masqués, Bar et al., (2001) ont pu montrer que l'activité corticale dans les régions visuelles ventro-temporales était linéairement corrélée avec les performances d'identification.

La relation entre l'activité corticale et l'identification consciente d'un stimulus a aussi été démontrée dans des paradigmes de rivalité binoculaire. Dans ce type de protocole, un stimulus A est présenté à un œil et un stimulus B est présenté à l'autre œil. La perception du sujet alterne entre le stimulus A et le stimulus B. Tong, Nakayama, Vaughan, & Kanwisher, (1998) ont présenté un visage à un œil et un bâtiment à l'autre œil tout en enregistrant les réponses BOLD (réponses dépendantes du niveau d'oxygène sanguin) en IRMf dans FFA et PPA. Alors que la stimulation visuelle restait constante (le stimulus ne changeait pas), la perception consciente du bâtiment vers le visage était accompagnée d'une augmentation de l'activité de FFA et d'une diminution de l'activité de PPA. L'inverse était observé quand la perception des sujets allait du visage au bâtiment.

Un autre type d'expérience dans lequel le stimulus ne change pas alors que la perception, elle, change est la présentation de stimuli ambigus. Le stimulus ambigu de Rubin (Figure 3) a ainsi été utilisé pour étudier la différence entre le niveau d'activité cérébrale et la perception consciente des sujets. Lorsque les participants indiquaient percevoir les deux visages, l'activité des régions sélectives aux visages augmentait par rapport à la condition où les participants rapportaient percevoir le vase (Andrews,

Schluppeck, Homfray, Matthews, & Blakemore, 2002; Hasson, Hendler, Ben Bashat, & Malach, 2001). Ces différents résultats confirment que l'activité des aires dédiées au traitement de stimuli spécifiques est corrélée positivement avec la perception consciente chez l'humain.

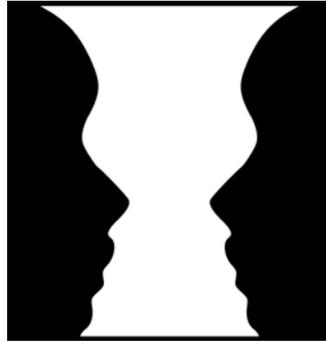


Figure 3. Vase d'Edgar Rubin.

1.1.2.3 Dissociation fonctionnelle des aires sélectives à certains stimuli

Avec l'utilisation de la TMS, différentes équipes ont pu montrer chez l'humain le lien entre augmentation du flux sanguin (réponses BOLD) et rôle fonctionnel. Par exemple, Pitcher, Charles, Devlin, Walsh, & Duchaine, (2009) ont testé des participants dans des tâches de discrimination de 3 types de stimuli : visages, corps et objets. Les participants devaient juger si le second stimulus était différent ou identique au premier. La stimulation pouvait être appliquée soit sur l'aire occipitale droite sélective aux visages (rOFA, Gauthier et al., 2000), soit sur l'aire extrastrée droite sélective aux corps (rEBA, Downing et al., 2001), soit sur l'aire occipitale latérale droite sélective aux objets (rLOC, Malach et al., 1995). Les résultats de cette étude montrent une triple dissociation entre les visages, les corps et les objets : les performances des participants étaient altérées seulement quand la stimulation était appliquée sur la région corticale sélective pour la catégorie de stimuli testés (par exemple, la TMS sur rOFA diminuait les performances de discrimination de visages mais pas celles pour les objets ou les corps). De la même façon, des résultats ont confirmé le lien entre LOC et la catégorisation d'objets et celui entre le sulcus occipital transverse et la catégorisation de scènes (Dilks, Julian, Paunov, & Kanwisher, 2013; Ganaden, Mullin, & Steeves, 2013; Mullin & Steeves, 2011).

1.1.3. Vers une représentation distribuée des catégories d'objets

1.1.3.1 Représentation des catégories d'objets au sein des POPULATIONS de neurones de IT

Les études en électrophysiologie mentionnées dans la partie précédente montrent que les neurones de IT peuvent être activés sélectivement par des objets complexes tout en étant tolérants à certaines transformations de ces objets (position, taille, orientation, éclairage). Cependant, des neurones sélectifs à un objet en particulier sont très rares. Lorsque de très larges sets d'images sont utilisés, la plupart des neurones de IT répondent à de nombreux objets (Brincat & Connor, 2004; Desimone et al., 1984; Freedman, Riesenhuber, Poggio, & Miller, 2006; Kreiman et al., 2006; Logothetis, Pauls, & Poggio, 1995; Rolls & Tovee, 1995; Rolls, 2000; Vogels, 1999b; Zoccolan, Kouh, Poggio, & DiCarlo, 2007).

Les travaux de Vogels, (1999b) suggèrent que l'activité des neurones de IT seule est insuffisante pour coder les catégories d'objets mais qu'au contraire, une population de ces neurones peut fournir suffisamment d'information pour catégoriser des images complexes. Dans son étude, deux singes ont été entraînés à catégoriser des images d'arbres ou d'objets complexes non-arbre (Vogels, 1999a). La plupart des neurones répondaient pour les arbres et les non-arbres, environ ¼ répondaient seulement pour les arbres (activité spécifique à une catégorie d'objets) et 1/10 répondaient exclusivement aux exemplaires utilisés pour l'entraînement (et pas aux nouveaux stimuli arbres). Les réponses de ces neurones étaient invariantes aux transformations des stimuli (position et taille), et diminuaient avec le degré de brouillage de l'image (scrambling), en accord avec les performances comportementales. D'après Vogels, les catégories d'objets seraient représentées par des patterns d'activité de populations de neurones. Au sein de cette population, chaque neurone répondrait à un set limité d'exemplaires de la catégorie. La catégorisation d'un nouveau stimulus serait alors basée sur (a) la similarité du profil d'activation de la population de neurones élicité par d'autres exemplaires de la catégorie apprise (b) la différence du profil d'activation produit par les exemplaires d'autres catégories.

Les neurones unitaires de IT semblent agir comme des éléments d'une population qui, prise dans son ensemble, sert de support à la reconnaissance d'objets. Cette idée incite à utiliser des approches de corrélation ou de classification des patterns d'activité cérébrale. Ces types d'analyse seront abordés dans la partie 4.

1.1.3.2 Un réseau d'aires distribuées à la base de la reconnaissance d'objets

Malgré l'observation d'aires cérébrales du cortex temporal ventral activées fortement pour des types de stimuli particuliers, il semble improbable que n'importe quel visage, outil, bâtiment, et autres catégories d'objets puissent avoir une région spécialisée ou un ensemble de neurones dédié à sa représentation. En effet, il y a plusieurs façons de catégoriser un visage ou un objet (par exemple, un visage peut être catégorisé comme étant féminin vs. masculin, âgé vs. jeune, vue de face vs. de profil, etc...).

L'idée d'un réseau distribué qui sous-tendrait les représentations des catégories d'objet est déjà présente en 1999. Dans leur étude, Ishai et al., (1999) ont observé trois régions distinctes du cortex temporal ventral qui répondaient préférentiellement aux visages, aux maisons et aux chaises. Cependant, en plus de ces trois régions séparées, ils ont aussi remarqué que chaque catégorie évoquait des réponses significatives dans d'autres régions cérébrales. En fait, chaque catégorie d'objet était associée à un pattern de réponse spécifique au sein du cortex temporal ventral. Ces résultats indiquent que la représentation d'un objet n'est pas restreinte à une région mais qu'elle serait le fruit d'une activité cérébrale distribuée. La voie ventrale ne serait donc pas une mosaïque de modules spécialisés dans la représentation de catégories d'objets mais représenterait l'information de manière continue et cohérente (Ishai, Ungerleider, Martin, & Haxby, 2000).

L'organisation modulaire ou distribuée des catégories d'objets fait débat plusieurs années (Haxby et al., 2001; Op de Beeck, Haushofer, & Kanwisher, 2008). La controverse portait notamment sur la question de savoir si les régions qui élicitent une réponse maximale pour une catégorie d'objet devraient être considérées comme des modules pour la représentation de cette catégorie ou si elles feraient partie d'un système de reconnaissance d'objets plus général (Avidan et al., 2002; Haxby et al., 2001; Spiridon & Kanwisher, 2002). En tout cas, l'idée d'un réseau distribué des différentes catégories d'objet a fait naître de nouvelles façons d'analyser les données en imagerie afin d'inclure beaucoup plus de régions corticales et de s'intéresser aussi à des patterns d'activités au sein de régions d'intérêt.

1.1.4. Nouvelles idées, nouvelles techniques, nouveaux résultats

Des techniques récentes d'analyses nous permettent de pouvoir interpréter l'activité cérébrale dans son ensemble. Au lieu de s'intéresser individuellement à chaque neurone ou voxel, l'activité d'un pattern de neurones ou voxels évoqué pour un stimulus est analysée afin de déchiffrer l'information présente dans ce pattern d'activité.

En IRMf, cette approche est appelée MVPA (multi-voxel pattern analysis). En prenant en compte l'activité cérébrale dans son ensemble (mesurée simultanément à plusieurs endroits), la sensibilité de la neuroimagerie chez l'humain est énormément augmentée. Les techniques de MVPA permettent de différencier l'activité de patterns évoqués dans deux conditions différentes, même si la moyenne de ces patterns d'activité prise séparément est la même et donc indissociable en utilisant seulement le niveau d'activité de cette région (Figure 4). L'existence de régions spécialisées dans le traitement d'un stimulus particulier (e.g. FFA) n'est plus nécessaire pour distinguer l'activité évoquée par différentes catégories d'objets.

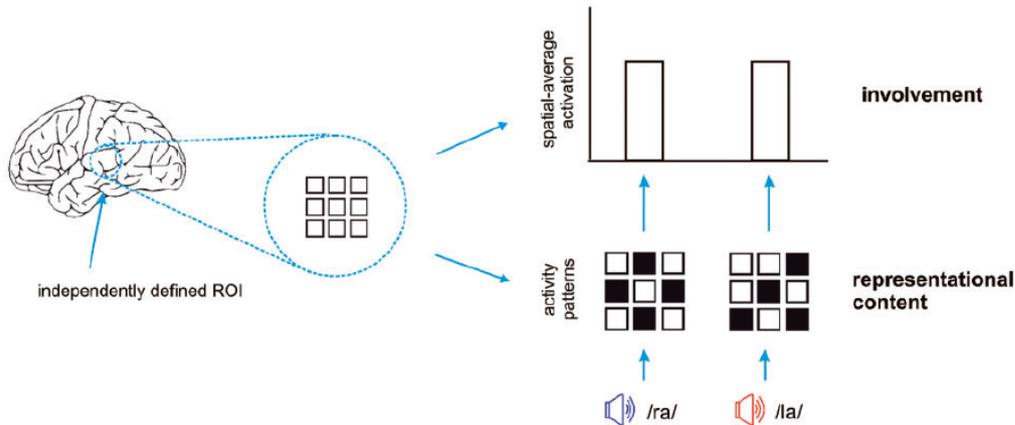


Figure 4. L'activation indique l'implication d'une région cérébrale alors que le pattern d'activité indique la représentation du contenu de l'information (les carrés noirs représentent des voxels actifs alors que les carrés blancs des voxels inactifs). Les sons /ra/ et /la/ évoquent des patterns d'activité différents (information différente) alors que la moyenne de l'activité de ces patterns est la même. (Mur, Bandettini, & Kriegeskorte, 2009)

1.1.4.1 Analyse de patterns d'activité

1.1.4.1.1 Matrices de dissimilarité

L'utilisation de techniques d'analyse multi-variées de patterns (MVPA) a été introduite par Haxby et al., (2001) et est maintenant très largement répandue. Les auteurs ont calculé la corrélation entre le pattern évoqué dans une condition et celui évoqué dans une autre condition. Si la catégorie d'un stimulus produit un pattern d'activité spécifique alors des observations indépendantes de patterns générés par cette catégorie devraient être plus similaires entre elles que les patterns produits par une autre catégorie.

Dans leur étude, chaque participant était scanné dans l'IRMf tandis qu'il/elle voyait 8 catégories de stimuli différentes (visages, chaises, bouteilles, maisons, etc.). Les données de chaque sujet étaient divisées en deux et le pattern d'activité du cortex ventro-temporal évoqué pour chaque catégorie était caractérisé pour chacune des deux moitiés (16 cartes d'activation au total). En corrélant la première moitié des patterns d'activité avec la seconde moitié des patterns d'activité pour chaque sujet séparément, les auteurs étaient capables de montrer que chaque catégorie était associée à un pattern d'activité fiable et distinct. Par exemple, la première moitié des patterns d'activité évoqués pour la catégorie « chaussures » correspondait à la deuxième moitié des patterns d'activité évoqués pour la même catégorie « chaussures » plus qu'elle ne correspondait aux patterns associés aux autres catégories d'objets (Figure 5).

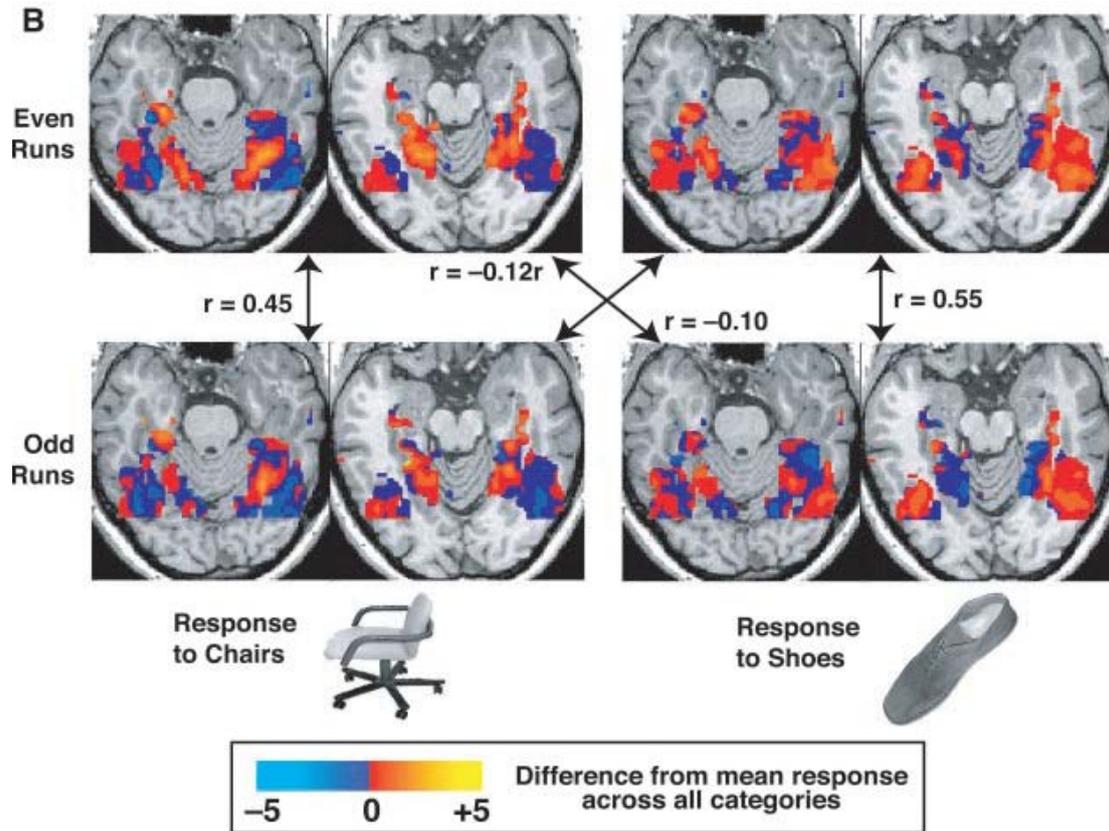


Figure 5. Comparaisons entre les patterns de réponse (corrélation r) évoqués pour la catégorie chaise et chaussure chez un sujet. (Haxby et al., 2001)

1.1.4.1.2 Classification de patterns d'activité

Les études comme celles de Haxby et al., (2001), moyennent l'activité cérébrale de plusieurs essais dans une condition pour la comparer (et la corrélér) à la moyenne de l'activité cérébrale d'une autre condition. Des méthodes plus sophistiquées de reconnaissance de patterns (LDA : Linear Discriminant Analysis, ou SVM : Support Vector Machines) permettent, du fait de leur plus grande sensibilité aux modifications du signal, une estimation presque instantanée de l'activité cérébrale mais aussi de pouvoir déterminer la catégorie d'un objet en un seul essai (revue Haynes & Rees, 2006; Mur et al., 2009; Norman, Polyn, Detre, & Haxby, 2006). Les bases de l'utilisation de ces méthodes de classifications statistiques proviennent des recherches dans le domaine des interfaces cerveau-machine (Andersen, Hwang, & Mulliken, 2010; Blankertz et al., 2003; Nicoletis, 2001; Wolpaw & McFarland, 2004).

Ce type d'analyse se déroule en deux temps (Figure 6). Premièrement, à partir de l'information de l'activité cérébrale (patterns d'activité de populations de neurones, en magnétoencéphalographie « MEG », en électroencéphalographie « EEG », ou en IRMf) et de l'information de la catégorie de l'objet correspondant à cette activité, un classifieur est entraîné à catégoriser un set de stimuli (dans l'exemple, à séparer les sons /ra/ des sons /la/). Cet entraînement est fait sur plusieurs essais et/ou plusieurs exemplaires de chaque catégorie de stimuli (l'activité n'est pas moyennée). A l'aide d'algorithmes mathématiques, le classifieur va tracer un hyperplan lui permettant de séparer au mieux les patterns correspondants à ces deux catégories de sons. Dans un second temps, les performances du classifieur sont évaluées essai par essai, à partir de l'activité cérébrale évoquée par des stimuli pour lesquels le classifieur n'a jamais été entraîné. On va ainsi calculer la proportion de sons qu'il catégorise correctement.

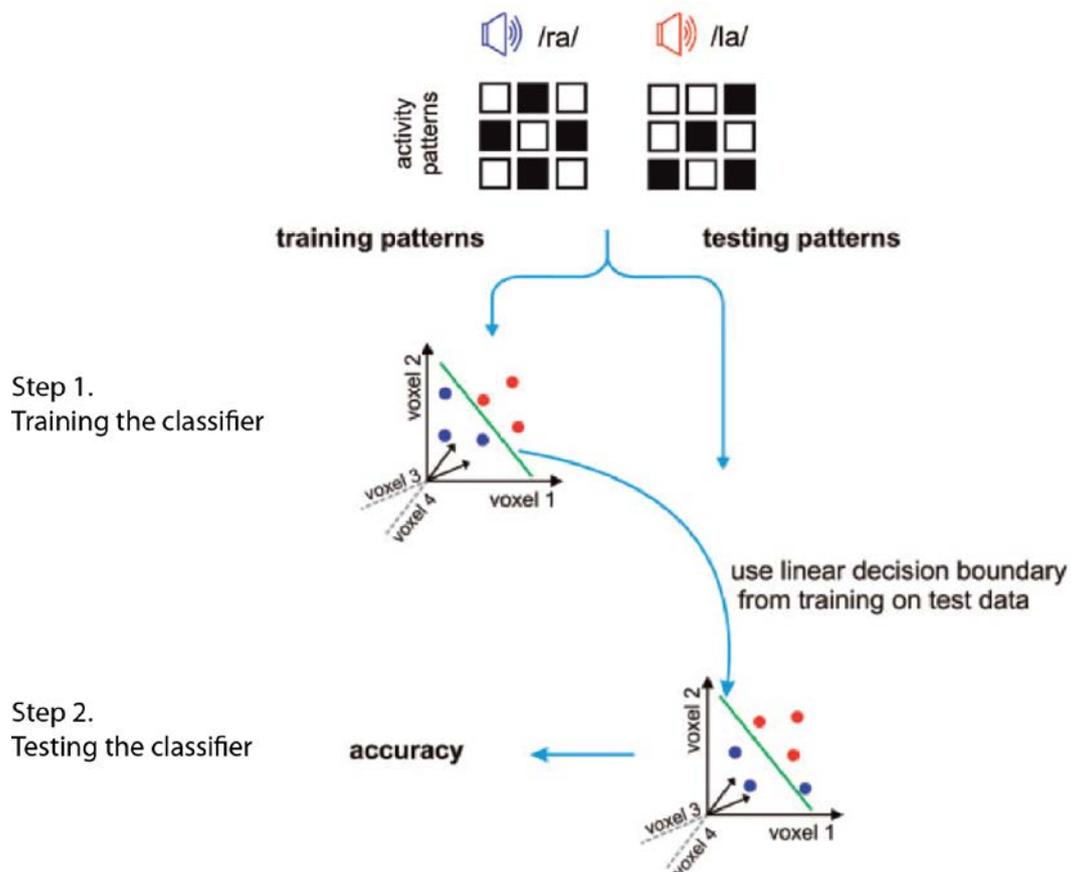


Figure 6. Analyse de patterns d'information. Les données sont d'abord séparées en deux sets. Un set sera utilisé dans une première étape pour entraîner le classifieur à catégoriser les patterns d'activité comme appartenant soit au son /ra/ soit au son /la/. Après avoir été entraîné, le classifieur est testé sur le deuxième set de patterns d'activité et les performances de classification sont calculées selon le nombre de patterns correctement catégorisés. Adapté de Mur et al., (2009).

Cependant, les méthodes de classification ont aussi des limites puisque ce qui est décodé reste encore mystérieux. En effet, plusieurs caractéristiques peuvent différencier des objets et chacune peut créer une différence dans le signal neuronal. Les analyses MVPA peuvent capter une caractéristique efficace pour distinguer deux catégories d'objets mais il n'y a aucune garantie que cette caractéristique soit pertinente cognitivement (Tong & Pratte, 2012). De plus, les analyses de classification posent comme principe l'existence d'un mécanisme de lecture (« read-out ») grâce auquel les informations utiles contenues dans les patterns d'informations sont extrapolées. Il n'est pas vraiment clair comment ce mécanisme pourrait être implémenté ni quelle(s) région(s) cérébrale(s) y participera(en)t.

1.1.4.2 Codage de plus en plus invariant dans la voie visuelle ventrale

En neurophysiologie, les techniques de classification ont été utilisées pour étudier le codage de la sélectivité et de l'invariance au niveau des populations de neurones de IT. Pour permettre la reconnaissance d'objets visuels, les patterns d'activités représentant les différentes catégories d'objets doivent être robustes aux modifications de la position de l'objet, de sa taille, du contexte, ou à son occlusion partielle. Par des simulations (N. Li, Cox, Zoccolan, & DiCarlo, 2009) ont montré que la propriété cruciale des réponses neuronales pour la reconnaissance d'objets n'était pas l'amplitude de réponse des neurones, mais la préservation pour chaque neurone de son ordre de préférence des objets (sélectivité). Cette propriété est présente dans les neurones de IT mais pas dans les neurones des aires visuelles plus précoces. Rust & DiCarlo, (2010) ont, eux, observé une augmentation de la tolérance aux modifications du stimulus (position, taille, contexte) entre V4 et IT. En général, au fur et à mesure que l'on avance dans les régions plus antérieures de la voie visuelle ventrale, les informations d'identification et de catégorisation des objets visuels deviennent de plus en plus précises et invariantes (Freiwald & Tsao, 2010; N. Liu et al., 2013).

Les résultats obtenus avec les techniques de classification utilisées en IRMf appuient les résultats obtenus en électrophysiologie chez le singe. A partir des patterns de réponses dans LOC, Eger, Ashburner, Haynes, Dolan, & Rees, (2008) ont montré qu'il était possible de déterminer au-dessus de la chance non seulement la catégorie des objets mais aussi les exemplaires de la même catégorie d'objets qu'un participant voyait (4 théières et 4 chaises). Plus important, ils ont aussi montré que la discrimination d'exemplaires était possible en utilisant des tailles et des orientations différentes pour

l'entraînement et le test du classifieur. Les informations contenues dans les patterns de réponses en IRMf représenteraient donc les objets de manière invariante et ne seraient pas limitées à la différence de forme globale des objets entre les catégories. L'identité d'un objet pouvait aussi être prédite par les réponses du cortex visuel primaire, cependant, une dissociation était observée avec LOC : dans le cortex visuel primaire, deux tailles différentes du même objet étaient beaucoup mieux discriminées que deux différents objets, tandis que l'opposé était vrai dans LOC.

Ces différents résultats suggèrent un codage distribué des représentations visuelles devenant de plus en plus invariant au fur et à mesure qu'on avance dans la voie ventrale. Le pattern d'activité produit à la sortie de la voie ventrale pourrait être à la base de la catégorisation et de l'identification des objets visuels.

1.1.4.3 Organisation distribuée et modulaire des catégories visuelles

De plus en plus de résultats en IRMf analysés avec les techniques de classifications étayent l'hypothèse que les catégories d'objets seraient représentées de façon distribuée dans le cortex occipito temporal ventral (Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Eger et al., 2008; Haxby et al., 2001; O'Toole, Jiang, Abdi, & Haxby, 2005). En effet dans ces études, les performances de classification augmentaient avec le nombre de voxels inclus dans l'analyse, en accord avec la notion que, même au sein d'une même catégorie, la représentation de la forme d'un objet n'est pas restreinte aux voxels les plus discriminants (comme proposé par les modèles distribués des représentations des catégories d'objets : Haxby et al., 2001; Ishai et al., 1999).

Cependant, on ne peut pas nier l'existence de certaines régions préférentiellement activées pour des stimuli particuliers, par exemple les visages. Le traitement des visages a d'abord été corrélé avec une augmentation d'activité dans FFA. Il a ensuite été montré que OFA était aussi sélective à ce type de stimulus. En fait, de nouvelles études ont identifiées un nombre croissant de régions sélectives aux visages et aux parties du corps dans le cortex visuel : deux régions dans le gyrus fusiforme (FFA-1 and FFA-2), une région dans le cortex temporal antérieur, et deux régions au niveau du sulcus temporal supérieur (Pinsk et al., 2009). Une étude en IRMf chez le primate non-humain a révélé jusqu'à six régions sélectives aux visages dans le lobe temporal, qui étaient similaires à ceux des humains en taille

relative et en nombre (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Tsao, Moeller, & Freiwald, 2008). En utilisant des méthodes de micro-stimulation, Moeller, Freiwald, & Tsao, (2008) ont aussi pu montrer que ces régions étaient interconnectées, créant un grand réseau sélectif aux visages.

Ces résultats indiquent que les informations sur la catégorie des objets sont distribuées au sein du cortex mais qu'il existe aussi des régions sélectives à certains stimuli qui peuvent former un réseau plus exclusif. La plupart des modèles de reconnaissance visuelle combinent maintenant une organisation distribuée et modulaire des représentations des objets (e.g. Op de Beeck et al., 2008; Weiner & Grill-Spector, 2013).

1.1.4.4 Un espace continu représentant les catégories d'objets

Les patterns de réponses pour différents objets visuels peuvent aussi être analysés en termes de degré de similarité entre ces objets afin de représenter ces similarités dans un espace en 2D ou 3D. Très tôt, Edelman, Grill-Spector, Kushnir, & Malach, (1998) utilisent ce type d'analyse et leurs résultats, ainsi que ceux de O'Toole et al., (2005), indiquent que la similarité perceptuelle entre les objets est reflétée dans les patterns de réponses à ces objets dans le cortex occipito-temporal.

Plus tard, en analysant la différence d'activité des populations de neurones de IT pour différentes catégories d'objets, Kiani, Esteky, Mirpour, & Tanaka, (2007) ont montré dans une étude ambitieuse que les catégories d'objets étaient organisées de façon hiérarchique (Figure 7). La différence entre l'activité des patterns de réponses était la plus grande entre les objets animés et les objets inanimés et au sein de ces deux grands groupes, des sous-catégories pouvaient être distinguées par exemple entre les catégories corps et visages appartenant toutes deux à la catégorie objet animé.

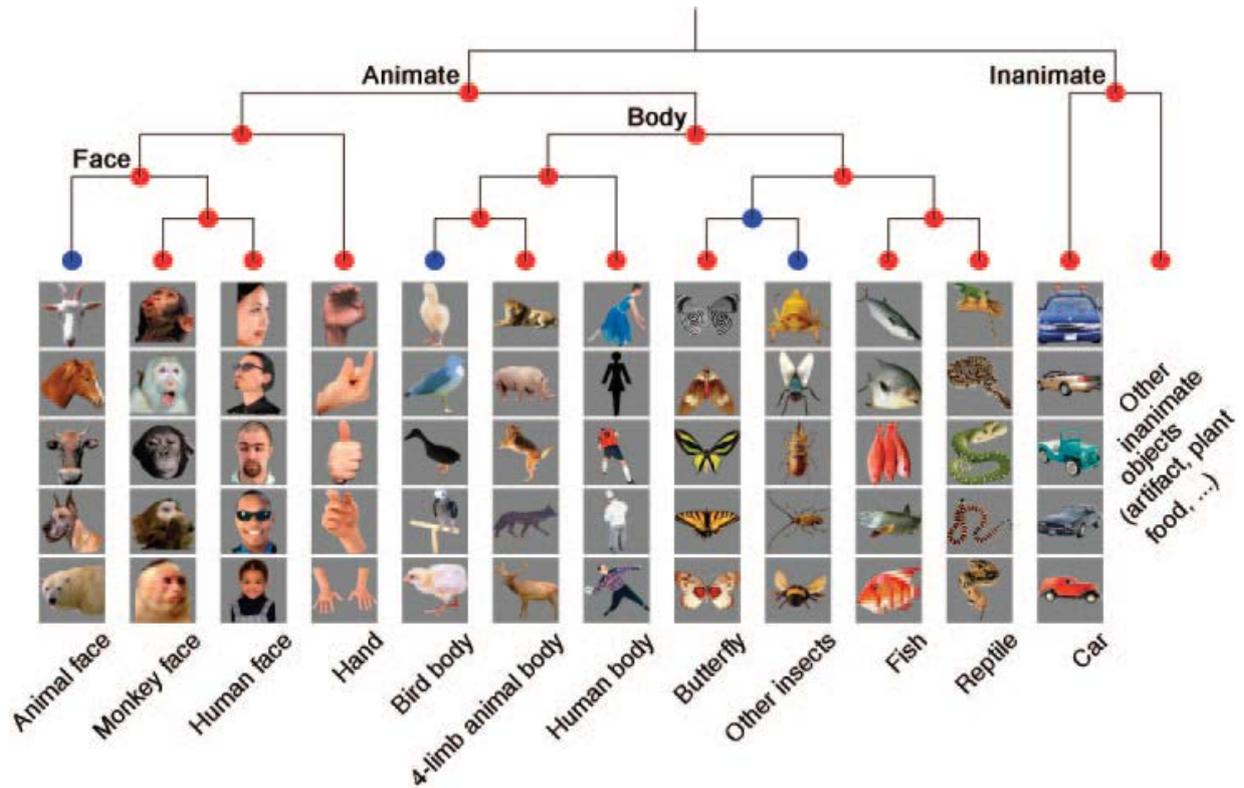


Figure 7. Reconstruction de la hiérarchie des catégories d'objets en fonction de la dissimilarité de leurs patterns de réponses neuronales dans IT (enregistrement chez le primate non-humain). (Kiani et al., 2007)

Ce même type d'analyse a été réalisé chez l'humain avec des enregistrements en IRMf. Kriegeskorte et al., (2008) ont présenté 92 images de différents objets naturels et ont évalué la similarité des patterns d'activité évoqués par ces stimuli. Les réponses des aires de IT (cortex temporal ventral) reconstruisaient la structure que nous avons des catégories d'objets, ce qui n'était pas le cas de V1 (Figure 8). Les objets animés vs. inanimés montraient des patterns d'activité très distincts dans IT chez l'humain, de la même façon que Kiani et al., (2007) l'avait montré avec l'analyse de patterns de réponses de neurones chez le singe. De plus, les patterns d'activité de IT chez l'humain permettaient une meilleure discrimination entre différents visages humains qu'entre différents visages de primates non humains, alors qu'une tendance dans le sens contraire était observée dans les données chez le singe.

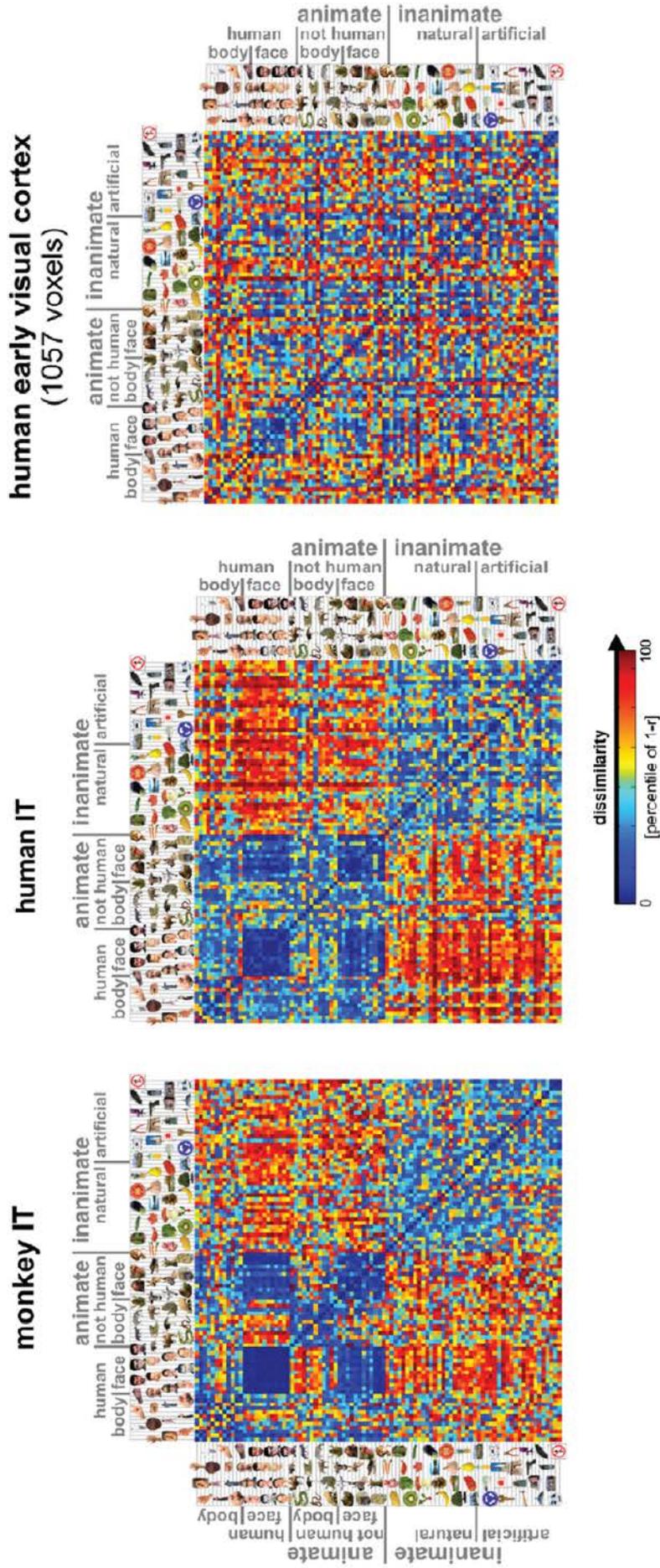


Figure 8. Matrices de dissimilarité dans IT chez le singe (Kiani et al., 2007), l'humain et dans les aires visuelles primaires chez l'humain. Pour chaque paire de stimuli un coefficient de corrélation entre les réponses évoquées pour les réponses évoquées pour les deux stimuli est calculé (r , corrélation de Pearson).

En fait, les patterns d'activité de IT pour différents objets sont fortement corrélés avec les jugements de similarité de ces même objets par des humains (Connolly et al., 2012; Mur et al., 2013). Cette corrélation était beaucoup moins forte calculée avec l'activité du cortex visuel primaire ou d'autres régions de la voie ventrale. Cependant, l'activité de IT chez l'humain était plus similaire avec celle de IT chez le singe qu'avec les jugements humains (Mur et al., 2013). Les auteurs proposent qu'au cours de son évolution, IT serait devenu un détecteur de traits visuels permettant de distinguer les objets animés des objets non-animés (ainsi que d'autres distinctions telles que corps vs. visages) alors que d'autres systèmes cérébraux serviraient à catégoriser les objets visuels différemment.

En considérant l'ensemble du cerveau, Carlson, Schrater, & He (2003) ont montré par des techniques de classification de patterns en IRMf que les voxels contribuant le plus à séparer les catégories chaises, visages et maisons étaient situées au sein de la voie visuelle ventrale. De plus, les informations contenues par ces voxels étaient largement indépendantes en termes d'activité mais aussi en termes de distribution spatiale (différents voxels étaient utilisés pour les catégories chaises, visages et maisons). Cette indépendance pourrait être la conséquence de l'utilisation de trois catégories d'objets relativement distinctes. En effet, Brants, Baeck, Wagemans, and Op de Beeck (2011) ont montré que des distinctions fines entre des objets proches conceptuellement (e.g. maison de campagne vs. gratte-ciel), sont aussi organisées dans le cerveau à une échelle beaucoup plus fine que de plus larges distinctions entre catégories (e.g. visage et maison).

Dans une autre étude analysant l'activité cérébrale dans son ensemble, Huth, Nishimoto, Vu, & Gallant, (2012) ont montré que la totalité du cerveau, et pas seulement la voie visuelle ventrale, contenait des informations sur les catégories d'objets. Dans leur étude des participants visionnaient des films alors que leur activité cérébrale était enregistrée en IRMf. Chaque image du film était étiquetée avec des noms de catégories lui correspondant (e.g. loup, carnivore, mammifère, etc.) et l'activité cérébrale était analysée en fonction de ces catégories et d'un modèle afin de construire un espace sémantique représentant ces catégories (1075 catégories d'objets et d'action ont été utilisées). Leurs résultats montrent que l'information sur les catégories d'objets est distribuée dans le cortex visuel et non-visuel. De plus, l'espace sémantique mais aussi l'organisation corticale de cet espace étaient partagés entre les cinq participants. Les auteurs proposent donc que les catégories d'objets seraient représentées dans un espace sémantique continu qui reflèterait la similarité conceptuelle des catégories

(idée aussi proposé par Connolly et al., 2012). Cependant les propriétés physiques des objets visuels pourraient aussi jouer un rôle non nul dans la représentation des catégories d'objets (voir Baldassi et al., 2013).

1.1.5. Résumé

Les nouvelles méthodes utilisées pour analyser l'activité cérébrale constituent des avancées prometteuses pour comprendre l'organisation cérébrale des représentations. Même si on ne connaît toujours pas exactement comment la voie visuelle ventrale arrive à coder des traits visuels en catégorie invariante, nous pouvons être assez certains qu'elle joue un rôle important dans l'implémentation de ces représentations invariantes. D'autres informations, sémantiques ou perceptuelles, sur les objets visuels sont aussi présentes dans d'autres aires corticales. Cependant, c'est bien dans la voie visuelle ventrale que les premières représentations visuelles sont établies. Qu'en est-il au niveau temporel ? A partir de quand avons-nous accès à la catégorie d'un objet ?

1.2. Quand ? Les catégories d'objets visuels sont accessibles en une centaine de millisecondes

Le temps nécessaire pour accéder à la catégorie d'un objet peut se poser de différentes façons. On peut mesurer le temps de présentation minimum pour qu'un objet puisse être traité en utilisant un paradigme de masquage (partie 1). On peut aussi mesurer le temps de réponse d'un sujet dans une tâche de catégorisation (partie 2). Une méthode encore plus précise pour déterminer le temps minimal pour traiter un stimulus visuel est d'enregistrer les réponses neuronales évoquées par un stimulus (partie 3).

1.2.1. Temps de présentation minimum pour avoir accès aux informations d'une scène visuelle

Afin d'étudier le temps de présentation minimum pour qu'un stimulus visuel soit traité, la seule présentation brève d'un stimulus n'est pas une méthode suffisante étant donné que l'activité des neurones visuels n'est pas limitée par la brièveté de la présentation du stimulus. Par exemple, des expériences ont montré que des neurones sélectifs aux visages dans le cortex temporal de primates répondent pendant 100 ms à un visage présenté pendant seulement 16 ms (Rolls & Tovée, 1994). Par contre, lorsqu'un stimulus visuel est suivi peu après par un autre stimulus (ou un masque), alors plus les deux images se suivent rapidement, plus la durée pendant laquelle les neurones répondent au stimulus test est réduite (Keysers, Xiao, Földiák, & Perrett, 2001; Kovács, Vogels, & Orban, 1995; Rolls, Tovée, Purcell, Stewart, & Azzopardi, 1994; Rolls & Tovée, 1994). Ce type de paradigme expérimental (appelé paradigme de masquage) est donc une méthode appropriée pour étudier le temps de présentation minimum nécessaire au traitement d'un objet visuel.

Dans un paradigme de masquage, la présentation brève d'une image (stimulus testé) est rapidement suivie de la présentation d'une seconde image (stimulus masque), ce qui va limiter le temps de traitement disponible pour traiter la première image. Une méthode similaire à ce protocole consiste à présenter rapidement une séquence de stimuli (RSVP : « rapid serial visual presentation »), où chaque stimulus devient le masque du précédent. Ce type de paradigme a été utilisé par Potter (1976). Les

participants devaient reconnaître un objet dans une série de 16 images présentées à 113, 167, 250 ou 333 ms par image. Il était demandé à un premier groupe de participants de presser un bouton dès l'apparition de l'image correspondant à la description donnée au début de la série (e.g. une rue avec des voitures). Un second groupe voyait la série de 16 images et devaient ensuite reconnaître les images qu'ils venaient de voir parmi de nouvelles images. Les participants étaient capables de détecter une image présentée à 113 ms (plus de 70% de réponses correctes) mais leurs performances étaient très faibles dans la tâche de reconnaissance (seulement 11% de réponses correctes). Pourtant, les performances de reconnaissance augmentaient avec l'augmentation du temps de présentation de chaque image. En fait, dans une expérience précédente Potter & Levy (1969) ont montré que lorsque les images étaient présentées pendant 500 ms, les participants arrivaient à se souvenir de 62% d'entre elles, et de 90% des images si elles étaient présentées pendant 2 sec. L'intervalle de temps disponible entre deux stimuli visuels devant être mémorisés semble être un facteur plus important dans les processus de mémorisation que le temps de présentation du stimulus. L'encodage en mémoire nécessiterait environ une seconde, si on considère que durant ce laps de temps aucune autre image n'ait aussi besoin d'être encodée (Intraub, 1979, 1980). Nous reviendrons sur le rôle de la mémoire dans le chapitre 4, pour le moment, on retiendra que d'après ces résultats, une scène naturelle peut être identifiée même présentée pendant seulement 113 ms avant un masque.

Keyser et al. (2001) ont pu montrer que les temps de présentations utilisés dans les expériences de Potter pouvaient être encore diminués tout en maintenant de bonnes performances de reconnaissance. En effet, même pour des images présentées pendant seulement 14 ms dans une séquence RSVP, les participants étaient capables de détecter la présence d'une image au-dessus du niveau de la chance. Cependant, les séquences étaient très courtes (seulement 7 images) et les mêmes stimuli étaient répétés plusieurs fois pendant l'expérience. Ces résultats ont néanmoins été répliqués dans un paradigme de masquage simple (présentation du stimulus test suivie par un masque dynamique) par Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe (2005). Dans cette étude, les participants devaient réaliser une tâche de catégorisation en go/no-go : ils devaient relever le doigt d'un bouton en moins d'une seconde si l'image flashée seulement 6 ms contenait un animal et garder le doigt appuyé si ce n'était pas le cas. Lorsque le stimulus était immédiatement suivi d'un masque, les participants n'étaient pas capables de réaliser cette tâche. Par contre, si le masque apparaissait 6 ms après la présentation de l'image (12 ms SOA : « stimulus onset asynchrony », c'est-à-dire 12 ms entre le début de la présentation de l'image et le

début de la présentation du masque), leurs performances étaient déjà au-dessus du niveau de la chance et continuaient d'augmenter jusqu'à atteindre un plateau pour des SOAs de 44-81 ms.

Dans cette expérience, les participants cherchaient la présence ou l'absence d'un animal. Il est possible que si les participants ne savent pas à l'avance ce qui va leur être demandé, et n'anticipent donc pas le type d'image qui va leur être présenté, leurs performances de reconnaissance soient beaucoup moins bonnes. C'est ce qu'avait observé Potter (1976) en comparant les performances des participants lorsque la description de l'image était faite avant ou après l'essai. Cependant, étant donné que dans une condition les participants ne devaient retenir que la description d'une image pour faire la tâche alors que dans la deuxième condition ils devaient maintenir en mémoire les 16 images, la différence des performances de reconnaissance pourraient être le résultat de la capacité limitée de la mémoire. Pour pallier ce problème, Evans, Horowitz, & Wolfe (2011) ont présenté à des participants une série d'images (présentées pendant 20 ms chacune) dont une était une scène naturelle et les autres des masques sans signification. Leurs résultats montrent que lorsque la catégorie cible était spécifiée après l'essai, les performances des participants étaient moins bonnes que dans le cas où la catégorie cible était spécifiée avant l'essai, mais restaient bien au-dessus du niveau de la chance.

Il est néanmoins possible que pour reconnaître une image contenue dans une séquence RSVP ou présentée très brièvement, les participants ne détectent qu'un trait caractéristique de la catégorie cible sans identifier l'image dans son ensemble (Evans & Treisman, 2005). Toujours en utilisant un paradigme RSVP, Potter, Wyble, Pandav, & Olejarczyk, (2010) ont pu montrer que les participants identifiaient bien l'image cible dans sa globalité. Cette fois, il était demandé aux participants de chercher deux images d'une catégorie spécifiée (par exemple deux oiseaux) et ils devaient ensuite donner l'identité spécifique de ces deux images (par exemple cygne et aigle). Les résultats montrent que la photo d'un objet peut être identifiée comme appartenant à une catégorie cible à temps pour qu'un autre objet puisse aussi être sélectionné, encodé et reporté.

Ces résultats sont très intéressants pour se rendre compte de la rapidité à laquelle les informations visuelles sont extraites et interprétées. Par contre, si ces expériences nous renseignent sur le temps de présentation minimum pour accéder aux informations d'une scène visuelle, elles ne nous informent pas sur le temps total requis pour le traitement de ces informations (VanRullen, 2011). Nous sommes capables de détecter la présence d'un oiseau au sein d'une série d'image si il est présenté pendant au

moins 44 ms mais nous n'avons pas d'information sur la durée des traitements sous-jacents (temps entre l'arrivée du premier photon sur la rétine jusqu'à la réponse motrice).

1.2.2. Temps de réponse minimum pour déterminer la catégorie d'un objet

Un des premiers articles qui explore le temps minimum pour déterminer la catégorie d'un objet est celui de Potter & Faulconer (1975). Les auteurs se sont aperçu que vérifier la catégorie d'une scène ou d'un objet (par exemple répondre « oui » ou « non » cette image est un animal) était considérablement plus rapide (environ 300 ms) que de nommer l'image. Mais c'est probablement l'article de Thorpe, Fize, & Marlot, (1996) qui a révolutionné le domaine de la reconnaissance visuelle. Dans cette expérience, des participants devaient réaliser une tâche de catégorisation go/no-go. Ils devaient lever le doigt si l'image était un animal et garder le doigt appuyé si ce n'était pas le cas. Les stimuli étaient des scènes visuelles naturelles complexes présentées pendant 20 ms. Etant donné la diversité des images, les participants n'avaient aucun a priori sur le type d'animal à chercher, sa position, sa taille, leur nombre, etc. Les participants ont réalisé cette tâche très facilement (94% de réponses correctes) pour un temps de réaction médian de 445 ms, et même de 382 ms pour les sujets les plus rapides. Encore plus impressionnant, si on calcule le temps de réponse minimal à partir duquel le nombre de réponses correctes est significativement plus élevé que le nombre de réponses incorrectes, alors il ne faut que 250 ms pour les humains (Delorme, Richard, & Fabre-Thorpe, 2000; Rousselet, Fabre-Thorpe, & Thorpe, 2002) et 180 ms pour les singes (Fabre-Thorpe et al., 1998) pour déterminer si un animal est présent ou non dans la scène visuelle.

Afin de tester la robustesse et la reproductibilité de ces résultats, des tâches de catégorisation ultra-rapides animal/non-animal ont été utilisées dans diverses conditions et les résultats de ces études montrent que l'utilisation d'images en couleur ou en niveaux de gris a peu d'impact sur les performances (Delorme et al., 2000), et que des images ayant seulement 10-12% de leur contraste initial sont catégorisées avec 70% de réponses correctes (Macé, Thorpe, & Fabre-Thorpe, 2005). De plus, ce type de catégorisation peut être réalisé au-dessus du niveau de la chance (60.5%) même à une excentricité de 70,5 degrés alors que l'image est présentée seulement pendant 28 ms à une position

pseudo-aléatoire (9 positions possibles sur l'axe horizontal) (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). Les très bonnes performances observées dans ces tâches de catégorisation ne sont pas dues à l'utilisation d'une catégorie spéciale (animal) puisqu'elles ont été répliquées avec des catégories artificielles (véhicules) (VanRullen & Thorpe, 2001a).

Pour aller encore plus loin, l'équipe de Thorpe a testé dans une tâche de catégorisation avec choix forcé des réponses encore plus rapides que les réponses manuelles : l'initiation d'une saccade. Deux images étaient flashées 20 ms de chaque côté d'une croix de fixation. Les participants devaient saccader le plus rapidement possible du côté de l'image contenant un animal. De manière impressionnante, les participants étaient capables de réaliser cette tâche avec 90% de réponses correctes en seulement 120 ms (Kirchner & Thorpe, 2006), c'est-à-dire environ 150 ms plus rapidement que pour des réponses manuelles (Figure 9). Des saccades encore plus rapides, peut-être pas vraiment contrôlées, peuvent être initiées en 100-110 ms pour les visages (avec un temps de réaction moyen de 140 ms) (Crouzet, Kirchner, & Thorpe, 2010). Le même paradigme a aussi été utilisé chez le singe dans une tâche de catégorisation animal/non-animal (Girard, Jouffrais, & Kirchner, 2008). Les résultats montrent que le temps minimum pour initier une saccade chez le singe était de 100 ms avec un pourcentage de bonnes réponses autour de 85%. Les auteurs de cette étude font remarquer que le début de l'activité d'un neurone de IT déclenchée par un stimulus visuel est aussi d'environ 100 ms. Ils proposent donc que ce type de catégorisation ultra-rapide pourrait être basé sur des traitements réalisés dans les aires extrastriées (peut être V4) et s'appuyer sur des indices de la forme globale de l'objet.

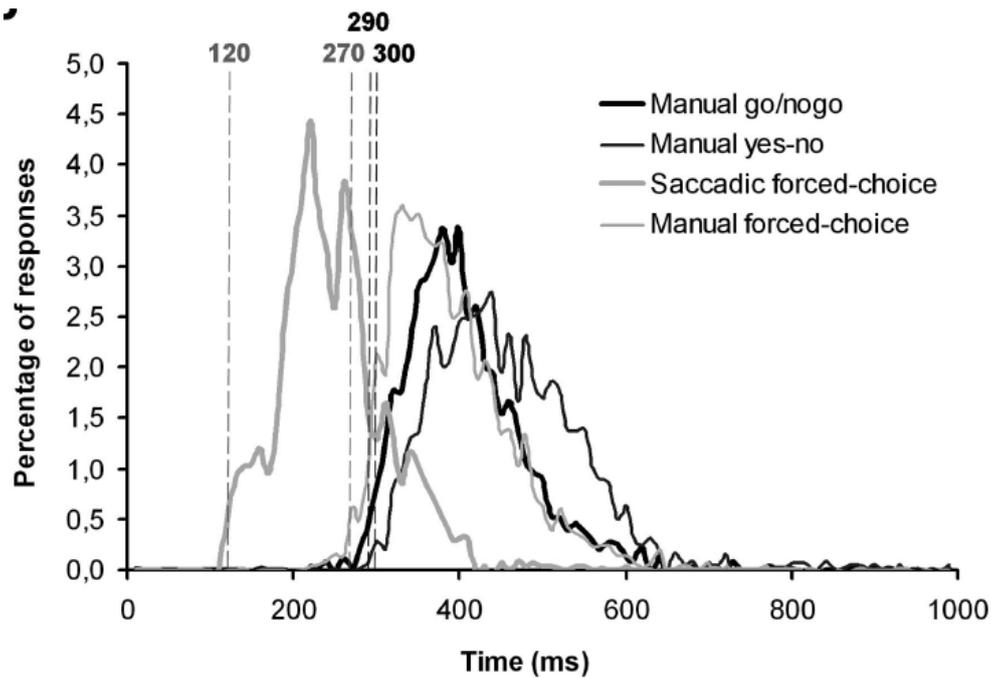


Figure 9. Distribution des temps de réaction dans une tâche de catégorisation animal/non-animal (paradigme go/no-go ou yes/no) et dans une tâche de choix forcé gauche/droite de localisation de l'animal (réponses en saccades ou manuelles). (Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, 2007)

Le temps de réaction pour catégoriser des images comprend : le temps pour que l'information visuelle soit transmise de la rétine au cortex visuel, puis traitée dans les différentes aires de la voie ventrale, avant l'intervention de mécanismes de prise de décision, et enfin l'envoi d'une commande motrice jusque dans la main (dans le cas des réponses manuelles). Ces deux dernières étapes se produisent après l'identification de l'image. L'accès aux catégories d'objets se fait donc dans des temps encore plus courts et on devrait pouvoir observer des différences neuronales entre les catégories très tôt après l'apparition d'un stimulus visuel.

1.2.3. Temps à partir duquel l'activité neuronale permet de différencier différentes catégories d'objets visuels

1.2.3.1 Activité de neurones unitaires

Les informations contenues dans l'activité des neurones de IT pourraient nous permettre de catégoriser très rapidement des stimuli visuels. En effet, certaines des études en enregistrement de neurones unitaires rapportant des activités neuronales sélectives pour des formes, des visages ou des catégories d'objets montrent aussi que ces activités apparaissent très tôt, dès 70-100 ms après le début de la présentation du stimulus (Z. Liu & Richmond, 2000; Perrett et al., 1982).

Des études ont essayé de quantifier la quantité d'information présente dans l'activité neuronale à l'aide d'analyses théoriques de l'information. Leurs résultats suggèrent que la majorité des informations encodées par l'activité neuronale semble être disponible dans une période très courte après le début des réponses des neurones de IT. Par exemple, Oram & Perrett (1992) ont montré qu'en utilisant une large population de neurones, il était possible de discriminer deux stimuli (différentes orientations d'un visage) à partir des 5 premières millisecondes du début de l'activité neuronale associée au stimulus visuel. Une autre étude (Tovée, Rolls, Treves, & Bellis, 1993) a analysé les réponses sélectives aux visages dans le cortex visuel temporal de macaques. Les résultats suggèrent que jusqu'à 87% de l'information contenue dans l'activité des neurones sur une période de 400 ms est aussi contenue dans les premières 50-100 ms de l'activité neuronale. De plus, environ 67% de l'information présente dans une fenêtre de 400 ms est disponible dès 100-120 ms après la présentation du stimulus.

En utilisant le paradigme de masquage, d'autres expériences ont pu montrer qu'une courte période d'activité des neurones est suffisante pour identifier un objet (Rolls et al., 1994; Rolls & Tovée, 1994). Ces études enregistraient l'activité de neurones de IT chez le singe lors de la présentation d'un visage pendant 16 ms. Sans masque, la réponse des neurones persistait pendant 200 à 300 ms, bien au-delà du temps de présentation du stimulus. Si un masque était présenté juste après le stimulus, alors plus le délai entre les deux stimuli était court (diminution du SOA), plus la durée pendant laquelle les neurones répondaient au stimulus test était réduite. Ainsi, lorsque le SOA était de 20 ms, les neurones sélectifs aux visages dans IT répondaient pendant 20 à 30 ms avant que leur activité ne soit interrompue par le masque. En utilisant le même protocole et les mêmes stimuli, les humains étaient capables d'identifier le visage qui leur avait été présenté parmi six autres visages avec 50% de réponses

correctes (leurs performances étaient de 97% avec un SOA de 40 ms). Ces résultats suggèrent que seulement 20 à 30 ms d'activité des neurones de IT sont suffisantes pour discriminer des objets visuels.

Ces résultats sont cohérents avec ceux de Keysers et al., (2001). Dans un paradigme RSVP, les auteurs présentaient à des singes macaques des scènes naturelles en utilisant différents temps de présentation (de 14 ms à 222 ms) et enregistraient simultanément dans IT la réponse de neurones sélectifs à des stimuli complexes (comme des visages). Les réponses neuronales des singes permettaient de discriminer un stimulus préféré d'autres stimuli vers 108 ms, c'est-à-dire dès que les neurones commençaient à répondre. Ils ont aussi pu montrer que la sélectivité de 65% de neurones était préservée même pour des présentations très rapides de 14 ms/image. La reconnaissance très rapide des objets visuels pourrait donc s'appuyer sur le début de l'activité de neurones sélectifs à certains stimuli (en particulier des visages).

1.2.3.2 Potentiels évoqués

Les résultats comportementaux obtenus par l'équipe de Thorpe montrant des temps de réaction très rapides dans une tâche de catégorisation animal/non-animal (chapitre 1.2.2), ont aussi été accompagnés de résultats enregistrés en EEG. Les potentiels évoqués (ERP) montrent une activité différentielle entre animaux (cibles) et les non-animaux (distracteurs) vers 150 ms après le début de la présentation du stimulus (Thorpe et al., 1996) (Figure 10). Cette signature neuronale a été retrouvée dans les différents articles cités dans la partie précédente mais aussi par d'autres équipes (Codispoti, Ferrari, Junghöfer, & Schupp, 2006; Johnson & Olshausen, 2003). Par ailleurs, cette activité différentielle est invariante à la position du stimulus et à la charge attentionnelle (Fize, Fabre-Thorpe, Richard, Doyon, & Thorpe, 2005). Des activités différentielles plus précoces ont aussi été observées (Johnson & Olshausen, 2003; VanRullen & Thorpe, 2001b) mais au contraire de celle décrite ici, elles semblent plutôt corrélées aux propriétés de l'image et pas à sa nature (cible ou distracteur).

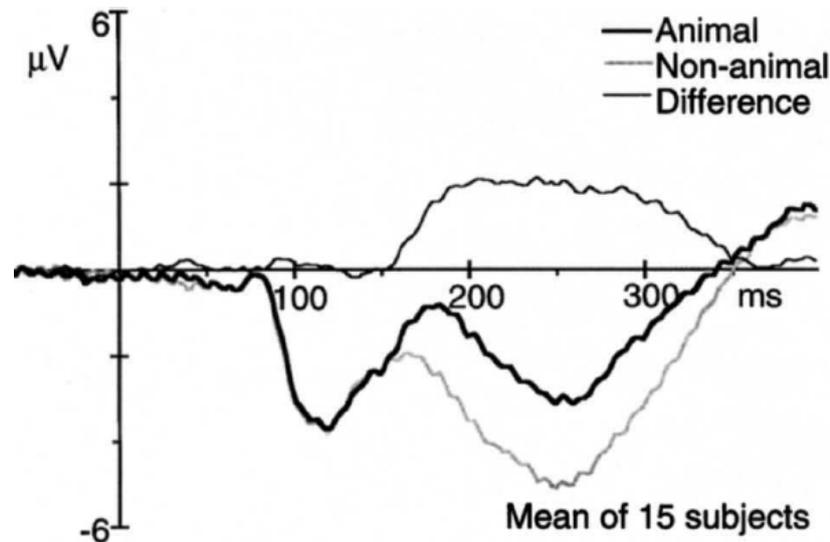


Figure 10. ERP (moyenne de sept électrodes frontales) des essais corrects pour les cibles (animal) et les distracteurs (non-animal). La différence entre ces deux conditions commence vers 150 ms après l'apparition du stimulus. (Thorpe et al., 1996)

L'idée que les ERPs permettent de différencier les différentes catégories d'objets visuels dès 150 ms après la présentation du stimulus est aussi appuyée par les résultats de Sadeh et al. (2011). Dans leur étude, les auteurs ont enregistré l'EEG de participants lors de la présentation de stimuli visuels (visages et corps) pour lesquels des aires spécialisées dans leur traitement ont été montrées en IRMf (Kanwisher & Yovel, 2006; Op de Beeck et al., 2008; Peelen & Downing, 2007) et par l'analyse des ERPs en EEG (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Thierry et al., 2006). En envoyant deux pulses de TMS à 60 et 100 ms après le début de la présentation du stimulus soit sur OFA soit sur EBA, ils ont pu montrer une double dissociation de la modulation de l'amplitude maximale des ERPs vers 150-200 ms entre le site stimulé et la catégorie du stimulus : la stimulation de OFA augmente significativement l'amplitude du pic des ERPs pour les visages mais pas pour les corps, tandis que la stimulation de EBA augmente significativement l'amplitude du pic des ERPs pour les corps mais pas celle des visages. Ainsi, la stimulation d'aires spécialisées dans le traitement de catégories d'objets visuels particulières influence spécifiquement le pic d'amplitude enregistré en EEG.

1.2.3.3 Classification de l'activité de populations de neurones

Chez le macaque, Hung, Kreiman, Poggio, & DiCarlo, (2005) ont pu montrer que l'activité d'une population d'une centaine de neurones des régions de IT contenait des informations précises et robustes sur l'identité et la catégorie des objets. Ces informations pouvaient être décodées dès 80-100 ms après la présentation du stimulus (Figure 11 à gauche). Les meilleures performances du classifieur, 70% de classifications correctes de la catégorie de l'objet, étaient obtenues vers 125 ms après l'apparition du stimulus. En augmentant le nombre de neurones inclus dans l'analyse, les performances de classification augmentaient mais elles restaient toujours meilleures pour déterminer la catégorie d'un objet que son identité (Figure 11 à droite). Les auteurs ont aussi montré que les informations contenues dans les patterns d'activité des populations de neurones de IT pouvaient être généralisées pour différentes positions et tailles des objets (invariance) ainsi que pour de nouveaux objets.

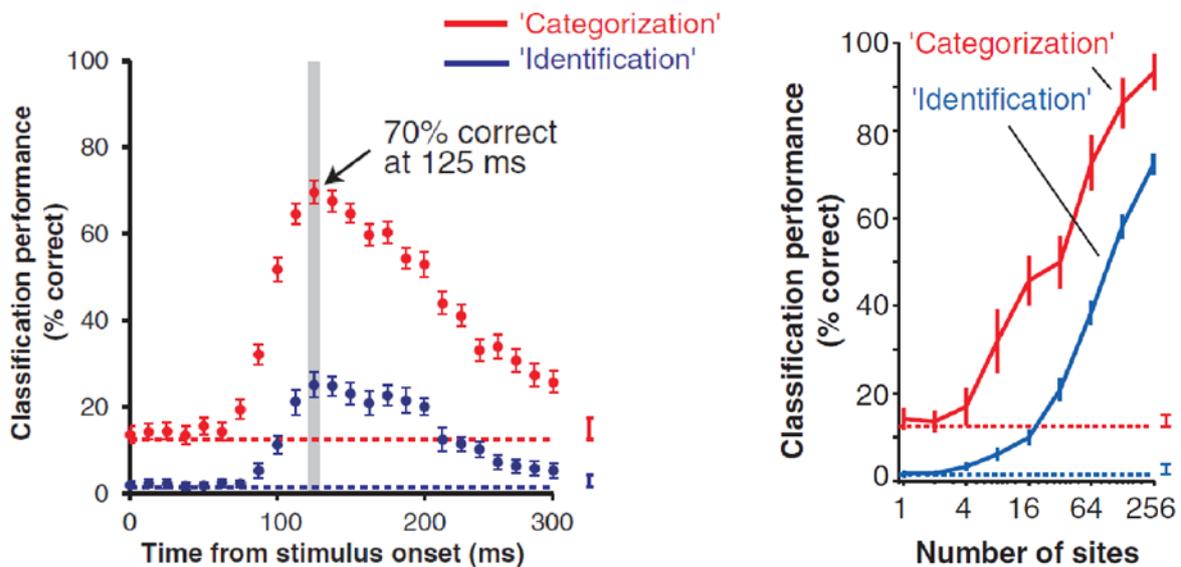


Figure 11. Analyse de l'information contenue dans les patterns d'activité de populations de neurones de IT. A gauche : performance de classification (bin de 12,5 ms) de la catégorie (en rouge) ou de l'identité (en bleu) du stimulus à partir de sa présentation. Les lignes en pointillés représentent le niveau de la chance, 12.5% pour la catégorisation et 1.3% pour l'identité. A droite : performance de classification de la catégorie ou de l'identité du stimulus en fonction du nombre de sites d'enregistrement inclus dans l'analyse. (Hung et al., 2005)

Les mêmes types d'analyse ont été utilisés à partir d'enregistrements intracérébraux (potentiel de champ local, LFP) des aires du cortex visuel chez onze patients épileptiques implantés pour des raisons médicales (H. Liu, Agam, Madsen, & Kreiman, 2009). Dès 100 ms après le début de la présentation d'un objet, des informations concernant sa catégorie (si c'était un visage, un animal, une chaise, un fruit ou un véhicule) était déjà présente. Les performances de classification étaient robustes aux modifications de taille, de profondeur et de rotation des objets.

Sans utiliser des enregistrements intracrâniens, le décours temporel du type d'informations contenues dans l'activité cérébrale peut être décodé grâce à l'utilisation de techniques de classification appliquées au signal EEG. L'information de la catégorie d'un objet contenue dans les patterns d'activité en EEG émerge dès 100 ms après la présentation du stimulus visuel avec les meilleures performances du classifieur observées vers 120-130 ms (Carlson, Tovar, Alink, & Kriegeskorte, 2013; B. Murphy et al., 2011; Simanova, van Gerven, Oostenveld, & Hagoort, 2010). Les possibles biais liés aux caractéristiques de bas niveau de l'image tels que la luminance, le niveau de détail, etc... n'expliquent pas les bonnes performances du classifieur (B. Murphy et al., 2011).

Une étude récente réalisée en MEG a pu isoler les différentes étapes du traitement visuel (Isik, Meyers, Leibo, & Poggio, 2013). En utilisant des techniques de classification, les auteurs ont montré que l'identité des objets pouvait être décodée dès 60 ms après le début de la présentation du stimulus. Les informations invariantes à la taille pouvaient elles être décodées vers 125 ms et celles invariantes à la position de l'objet plus tard vers 150 ms. De plus, les auteurs ont pu observer que ces informations invariantes émergeaient plus rapidement pour des petites modifications de l'objet que de plus grandes transformations. L'analyse de la localisation de l'activité cérébrale utilisée par le classifieur montrait que les informations les plus rapidement décodées étaient principalement basées sur l'activité des régions occipitales postérieures puis se déplaçaient le long du lobe temporal avec le développement de l'invariance.

1.2.4. Résumé

Les résultats de ces différentes études suggèrent que la voie visuelle ventrale construit en environ 150 ms une représentation invariante de la catégorie des objets. Ceci implique un traitement ultra-rapide

des informations visuelles provenant de la rétine. En fait, une première vague feed-forward (flux d'activité neuronale vers l'avant sans retour ou feed-back vers des aires inférieures) de l'information visuelle pourrait être suffisante pour réaliser des tâches de catégorisation telle que animal/non-animal. Dans le prochain chapitre, on exposera différents modèles pouvant expliquer la reconnaissance visuelle.

1.3. Comment ? Modèles du traitement de l'information visuelle

A partir des résultats obtenus au niveau comportemental et neurophysiologique sur l'organisation et la vitesse de traitement des catégories d'objets, différents modèles ont vu le jour. Je n'en présenterai ici que trois qui me semblent les plus importants et pertinents dans le cadre de cette thèse. Le premier modèle (chapitre 1.3.1) essaie plutôt d'expliquer l'organisation cérébrale des catégories d'objets en fonction de leurs caractéristiques sémantiques (e.g. animal qui peut voler). Les deux autres modèles (chapitre 1.3.2) essaient d'expliquer comment des stimuli visuels peuvent être catégorisés en très peu de temps grâce à un traitement majoritairement feed-forward de l'information.

1.3.1. Organisation hiérarchique des catégories d'objets

A partir de la nature hiérarchique des concepts (des catégories spécifiques « une mésange » aux catégories plus générales « un animal »), Collins & Quillian, (1969) ont proposé que les caractéristiques s'appliquant à tous les membres d'une catégorie générale (e.g. tous les membres de la catégorie animal ont des yeux) pourraient être stockées une seule fois à ce niveau et les autres caractéristiques s'appliquant à certains animaux mais pas à d'autres pourraient être stockées à un niveau inférieur (Figure 12). Le modèle prédit que les participants seraient plus rapides pour vérifier les propriétés spécifiques d'un objet que pour vérifier les propriétés partagées par tous les membres de la catégorie générale à laquelle il appartient. En effet, étant donné que les caractéristiques spécifiques sont directement stockées avec le concept mais que les caractéristiques générales sont stockées plus loin dans la hiérarchie, elles devraient être accessibles plus rapidement. Ce modèle se heurte donc aux vitesses de catégorisations très rapides obtenues dans les tâches de catégorisation animal/non-animal.

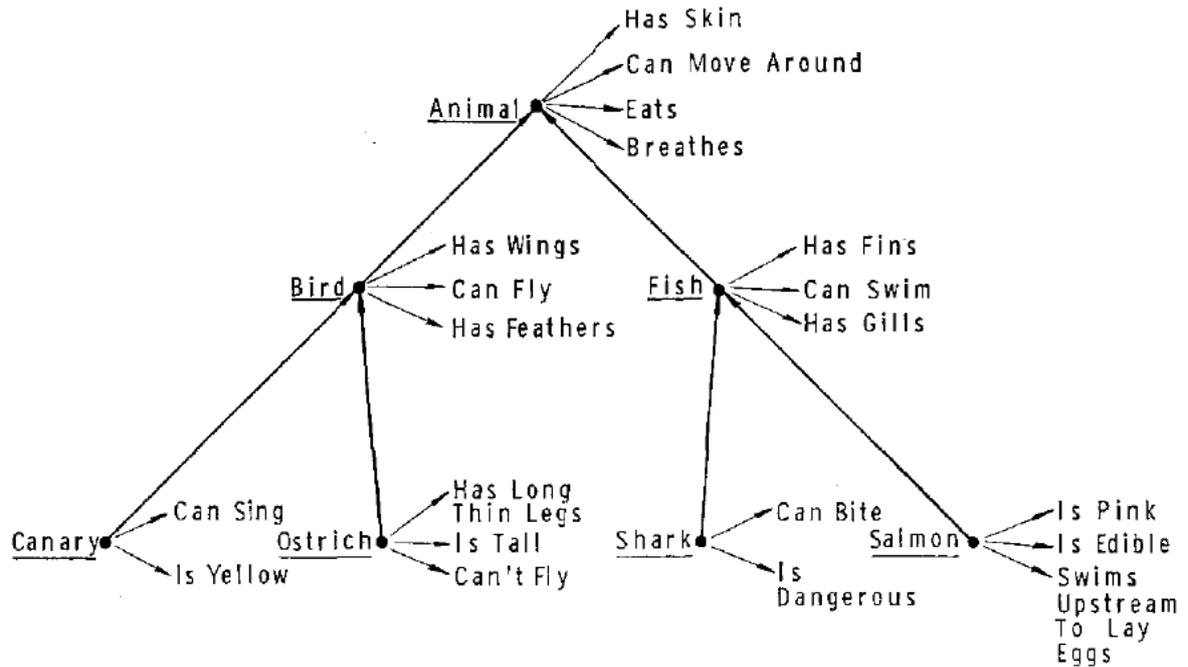


Figure 12. Illustration de la structure hypothétique des représentations des objets et de leurs caractéristiques associées. (Collins & Quillian, 1969)

A la suite de ce modèle hiérarchique, les modèles PDP (parallel distributed processing, traitement distribué et en parallèle) ont été développés. Ces modèles sont un sous-ensemble de modèles de réseaux de neurones artificiels qui utilisent des représentations distribuées : un stimulus est représenté par la participation simultanée de plusieurs unités (aboutissant donc à un pattern d'activité), ces unités pouvant aussi être utilisées pour représenter un autre stimulus. Les processus mentaux résultent alors d'interactions simples entre des unités similaires à des neurones et c'est le poids de leurs connexions qui change suivant le stimulus à traiter.

Le modèle d'abord introduit par Rumelhart & Todd, (1993) a une structure feed-forward, c'est-à-dire que l'activation se propage vers l'avant. McClelland, McNaughton, & O'Reilly, (1995) ont ensuite introduit l'idée d'un système d'apprentissage reposant sur deux modules complémentaires : un système d'apprentissage de connaissances sémantiques lent (impliquant des processus de mémoire), associé à un second système d'apprentissage rapide (rappelant ce qui se produit dans la voie visuelle ventrale).

Quand le réseau est initialisé, les patterns d'activation pour différents objets sont tous identiques. Le modèle est ensuite entraîné sur différents stimuli à l'aide des caractéristiques contenues dans la hiérarchie de Collins & Quillian, (1969) (voir Figure 12). Progressivement, les patterns d'activation deviennent de plus en plus différenciés, avec une progression du général au spécifique (Figure 13). Cette différenciation résulte de l'occurrence simultanée de caractéristiques partagées par un groupe d'objets les différenciant d'autres catégories (par exemple tous les animaux, au contraire des plantes, ont des yeux, ils mangent, etc.). La différenciation entre plantes et animaux arrive en premier parce que tous les objets appartiennent à l'une ou l'autre catégorie. La différenciation des objets individuels est plus tardive puisqu'ils sont différenciés par quelques caractéristiques qui ne covarient pas avec d'autres propriétés.

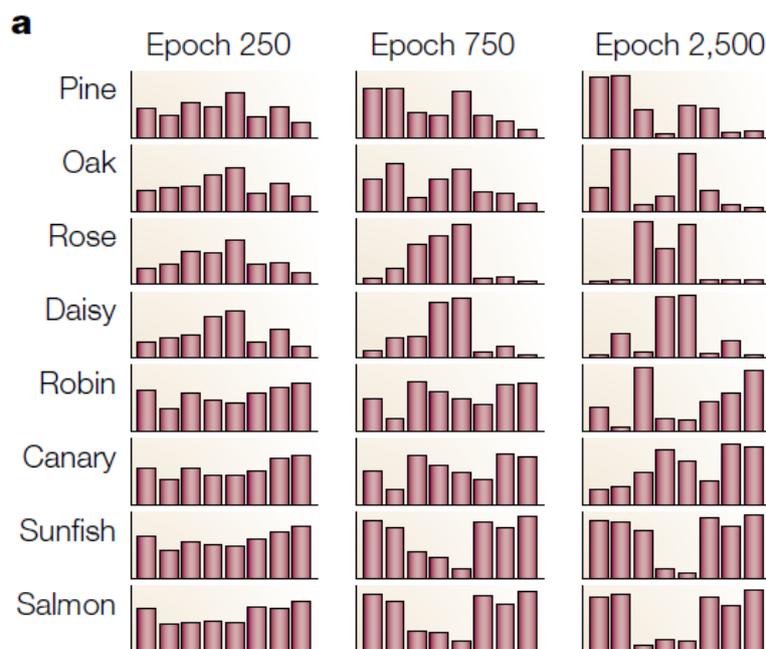


Figure 13. Différenciation des représentations conceptuelles au cours du temps dans le modèle PDP. Représentation de 8 patterns d'activité pour 8 objets à aux temps 250, 750 et 2500 depuis le début de l'entraînement. Au début de l'apprentissage, les patterns d'activité sont les mêmes pour tous les objets. La première différence apparaît entre les plantes et les animaux (t=250) puis les patterns deviennent plus différenciés pour les concepts intermédiaires (t=750 : arbre-fleur, oiseau-poisson) et enfin pour les concepts individuels (t=2500). Au final, tous les concepts ont des patterns d'activité différents mais ils gardent une organisation hiérarchique en fonction de leur similarité. (McClelland & Rogers, 2003)

Ce type de réseau reproduit la représentation hiérarchique des concepts de Collins & Quillian, (1969). Cependant, au lieu que les liens entre les concepts soient explicites (e.g. « un oiseau est un animal »), ils sont représentés au sein des similarités et des différences des patterns d'activité qu'ils évoquent. Ce modèle explique aussi comment l'organisation hiérarchique des patterns d'activité observés dans IT (Kiani et al., 2007; Kriegeskorte et al., 2008) et décrite dans le chapitre 1.1.4.4, pourrait être mise en place à partir de traits perceptuels communs à certaines catégories d'objets et différents d'autres catégories d'objets.

1.3.2. Une première vague feed-forward pour expliquer la catégorisation rapide

1.3.2.1 Latences des étapes de la catégorisation visuelle

Les humains sont capables de catégoriser des scènes visuelles naturelles très complexes en seulement 250 ms (Rousselet et al., 2002; Thorpe et al., 1996), et en seulement 180 ms pour les primates non humains (Fabre-Thorpe et al., 1998). En prenant en compte le temps nécessaire pour réaliser une réponse motrice, ces résultats suggèrent que l'image est traitée en moins de 200 ms. En fait, la différentielle d'activité en ERP entre les images animal et non-animal est observée 150 ms après l'apparition du stimulus. Ces résultats contraignent énormément le temps alloué à chaque étape du traitement visuel : de la rétine au CGL jusqu'aux aires de la voie ventrale puis au cortex préfrontal, au cortex moteur et enfin à la réponse manuelle (Figure 14). Des boucles de feed-back entre ces différentes étapes ne permettraient pas des réponses motrices si rapides. Thorpe & Fabre-Thorpe, (2001) ont ainsi proposé que les catégorisations d'objets ultra-rapides soient réalisées avec seulement une première vague feed-forward de l'information visuelle.

De nombreux résultats en électrophysiologie déjà évoqués dans le chapitre 1.2.3 (e.g. Hung et al., 2005; H. Liu et al., 2009) étayaient cette hypothèse. Par conséquent, il est maintenant largement admis que la catégorisation rapide d'objets soit soutenue par une cascade feed-forward de l'information visuelle (DiCarlo et al., 2012).

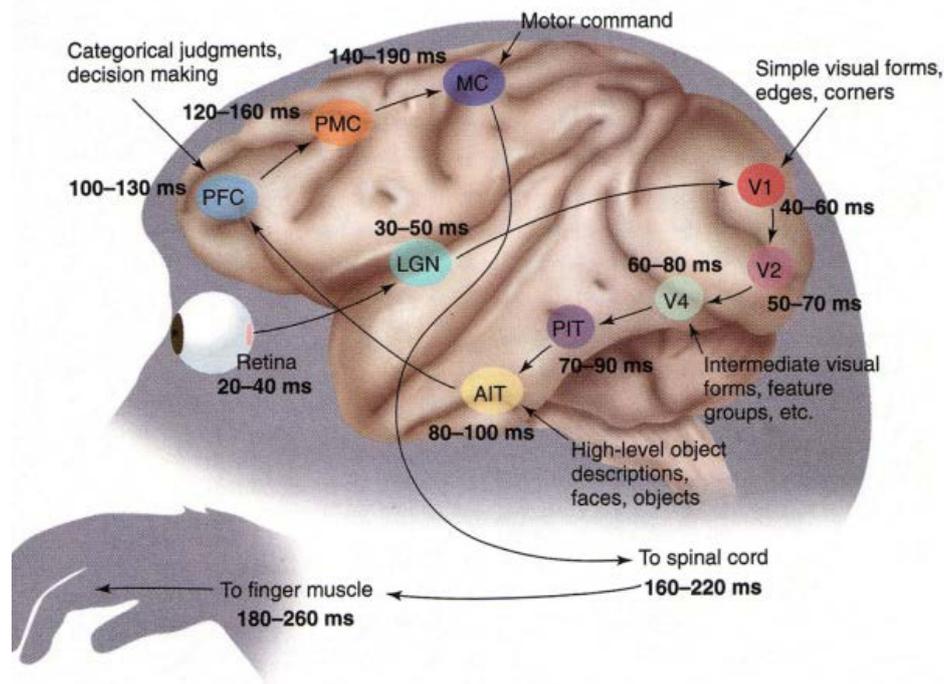


Figure 14. Schéma du chemin possible de l'information visuelle entre la rétine et les muscles de la main pendant une tâche de catégorisation (représentation pour un cerveau de primate non-humain). Les premières latences de réponses neuronales pour un stimulus flashé ainsi que les latences moyennes sont indiquées pour chaque étape de traitement. (Thorpe & Fabre-Thorpe, 2001)

From input to output. Monkeys can categorize complex visual stimuli very quickly, with reaction times that average 250 to 260 ms but that can be as short as 180 ms. Depicted is a plausible route between the retina and the muscles of the hand during a categorization task. Information from the retina is relayed by the lateral geniculate nucleus of the thalamus (LGN) before reaching V1, the primary visual cortex. From there, processing continues in areas V2 and V4 of the ventral visual pathway before reaching visual areas in the posterior and anterior inferior temporal cortex (PIT and AIT), which contain neurons that respond specifically to certain objects. The inferior temporal cortex projects to a variety of areas, including the prefrontal cortex (PFC), which contains the visually responsive neurons that categorize objects (7). To reach the muscles in the hand, signals probably need to pass via the premotor cortex (PMC) and primary motor cortex (MC) before reaching the motor neurons of the spinal cord. For each processing stage, two numbers (in milliseconds) are given: The first is an estimate of the latency of the earliest neuronal responses to a flashed stimulus, whereas the second provides a more typical average latency.

3.2.2. Modèles computationnels

Les temps de réponses extrêmement rapides observés par Thorpe et al. (1996) dans une tâche de catégorisation animal/non-animal ont aussi été à la base de modèles computationnels feed-forward. L'idée principale de ces modèles était de construire une représentation complexe et invariante des objets visuels au cours d'étapes successives tout en restant aussi proches que possible de l'anatomie et de la physiologie du cortex visuel. Par exemple, le modèle de VanRullen, Gautrais, Delorme, & Thorpe (1998) s'appuie sur le fait que les neurones d'une même couche du cortex déchargent à des moments différents avec le neurone le plus rapide étant aussi celui qui est le plus activé (Thorpe, 1990). Ainsi les auteurs ont proposé qu'un seul spike par neurone serait suffisant pour traiter un visage. Plusieurs autres modèles ont été développés (e.g. Delorme & Thorpe, 2001; Riesenhuber & Poggio, 1999; Serre, Wolf, & Poggio, 2005; Thorpe, Delorme, & VanRullen, 2001; VanRullen & Thorpe, 2002), le modèle le plus connu étant celui de Serre, Oliva, & Poggio (2007).

Ce dernier modèle intègre une large partie des connaissances sur les propriétés des neurones au sein de la voie visuelle ventrale (Figure 15). Il utilise un réseau d'unités pouvant être considérées comme des neurones artificiels avec celles de la première couche ayant des propriétés similaires à celles observées pour les neurones de V1 (e.g. sélectivité à des traits orientés). Deux types d'unités sont ensuite utilisés, chacune ayant des fonctions différentes. Les unités simples augmentent la complexité de la représentation visuelle en regroupant plusieurs unités afférentes avec des sélectivités pour différents traits de l'image. Les unités complexes augmentent la tolérance de la représentation en regroupant différentes unités ayant la même sélectivité mais avec de légères différences de positions et de tailles.

La première étape du modèle consiste en un apprentissage non supervisé des unités de V1 à IT. Ces unités vont s'ajuster aux régularités statistiques de l'environnement et devenir sélectives à certaines caractéristiques. Après cette étape, les unités au niveau le plus élevé du modèle (correspondant peut être aux unités du cortex préfrontal), sont entraînées sur un set d'exemples labellisés en fonction d'une tâche donnée pour que le modèle final puisse apprendre une tâche de catégorisation (apprentissage supervisé).

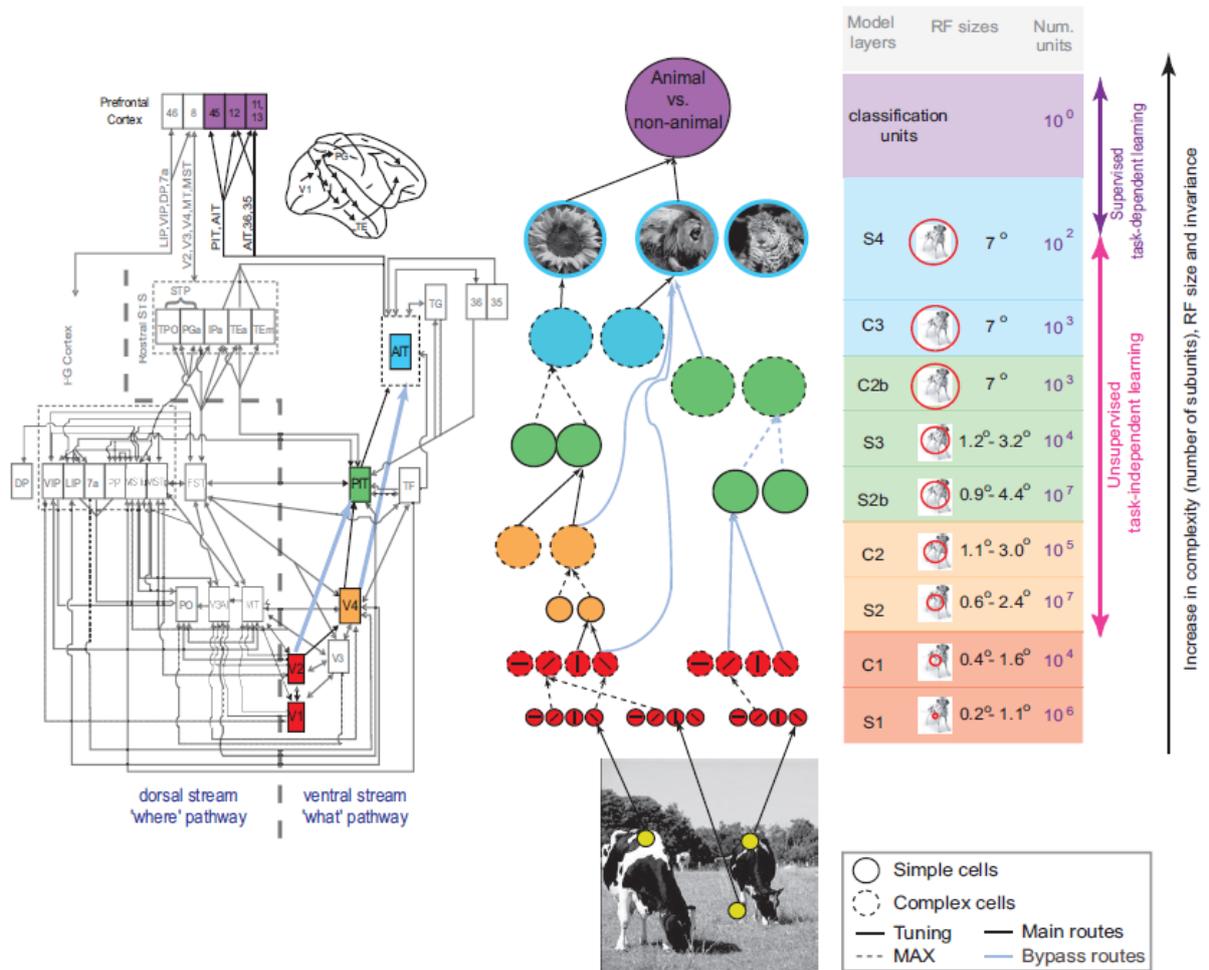


Figure 15. Représentation des étapes de traitement des informations visuelles (à gauche) avec leur correspondance (à droite) dans le modèle feed-forward de Serre, Oliva, et al., (2007). Voir le texte pour plus de détails.

Ce type de modèle est capable de reconnaître des objets du monde réel malgré leur modification d'apparence (éclairage, forme, ou avec seulement une partie visible). Ses performances dans une tâche de catégorisation animal/non-animal sont similaires à celles des humains (Crouzet & Serre, 2011; Serre, Oliva, et al., 2007; Serre, Kreiman, et al., 2007). Cependant, il ne tient pas compte des connexions feed-back qui pourraient jouer un rôle important dans la perception au-delà des 100-200 premières millisecondes. De plus, ce modèle permet seulement de réaliser des tâches de catégorisation, il ne permet pas d'interpréter une scène visuelle.

1.3.3. Résumé

Les modèles sont un travail de synthèse des connaissances difficile mais sans nul doute vraiment utile pour faire avancer les connaissances. L'utilisation de modèles permet de simuler et ainsi de tester différentes hypothèses quant au fonctionnement du système cérébral. Ils permettent aussi de générer de nouvelles hypothèses sur l'organisation et le traitement de l'information visuelle. Ces nouvelles hypothèses peuvent être testées dans de nouvelles tâches comportementales et les résultats du modèle ainsi que les résultats comportementaux peuvent être comparés.

1.4. Résumé de l'introduction

La difficulté de la reconnaissance visuelle est de pouvoir atteindre une grande sélectivité en maintenant une robustesse aux transformations des objets en quelques centaines de millisecondes. La représentation des catégories d'objets visuels se ferait de manière distribuée et deviendrait de plus en plus invariante à mesure qu'on progresse dans la voie ventrale. Par ailleurs, il semblerait que les objets visuels soient représentés dans un espace continu. En effet, la similarité entre les patterns de réponse dans IT est corrélée avec la similarité perceptuelle des stimuli. Les catégories d'objets sont de fait représentées de manière hiérarchique avec une majeure distinction entre objets animés et non-animés. La représentation des objets pourrait donc être basée sur des caractéristiques partagées ou distinctives d'autres objets.

De plus, ces représentations sont construites très rapidement. Les humains sont capables de catégoriser des images naturelles complexes en seulement 250 ms pour des réponses manuelles et 120 ms pour des réponses saccadiques. En fait, l'information de la catégorie de l'objet est déjà présente dans IT dès 100-150 ms après l'apparition du stimulus. L'ensemble de ces résultats suggèrent un traitement largement feed-forward de l'information visuelle provenant de la rétine.



Chapitre 2
Rapidité



2. Rapidité d'accès aux différents niveaux de catégories

Les objets peuvent être catégorisés de diverses façons, par exemple en fonction de leur couleur, leur taille, leur utilisation etc. Cependant, ces types de catégorisations font toujours appel à la même représentation d'un objet (même si des outils peuvent être catégorisés suivant leur couleur, ils font tous appel à la représentation 'outil'). De plus, il a été mentionné dans l'introduction que les représentations des objets deviennent invariantes à différentes modifications de taille, couleur, etc. au fur et à mesure de leur traitement. On ne s'intéresse ici pas aux catégorisations mettant en jeu les mêmes représentations mais plutôt aux différentes représentations d'un même objet. Par exemple, un objet (e.g. oiseau) peut être catégorisé à un niveau très détaillé (e.g. rapace), ou à un niveau plus abstrait (e.g. animal).

Dans le chapitre 2.1, on définira d'abord les différents niveaux de catégorisation (ou représentation) des objets. L'objectif sera ensuite d'essayer de déterminer dans l'article 1 le temps nécessaire pour accéder à différents niveaux d'abstraction : avons-nous accès aux représentations abstraites (e.g. animal) ou aux représentations plus détaillées (e.g. oiseau, rapace) en premier ? Les résultats de cet article sont discutés dans le chapitre 2.3.

2.1. Un objet, plusieurs catégories

Les humains catégorisent systématiquement les objets de leur environnement (un animal, un oiseau, une voiture, ...). Rosch, Mervis, Gray, Johnson, & Braem, (1976) ont montré que les catégories créées par l'homme seraient basées sur la fréquence de certaines combinaisons de caractéristiques naturelles. En effet, certaines combinaisons sont plus probables que d'autres. Par exemple, des attributs comme 'plumes' et 'ailes' cooccurrent souvent tandis que ce n'est pas le cas pour des combinaisons comme 'plumes' et 'roues'. La fréquence de cooccurrence de ces attributs serait à la base de la classification des objets. Dans cette étude, les auteurs définissent trois principales catégories d'objets : les catégories abstraites dites superordonnées (e.g. animal), les catégories basiques (e.g. oiseau) et les catégories plus restreintes dites subordonnées (e.g. passereau) (Figure 16). Nous utiliserons par la suite cette définition.

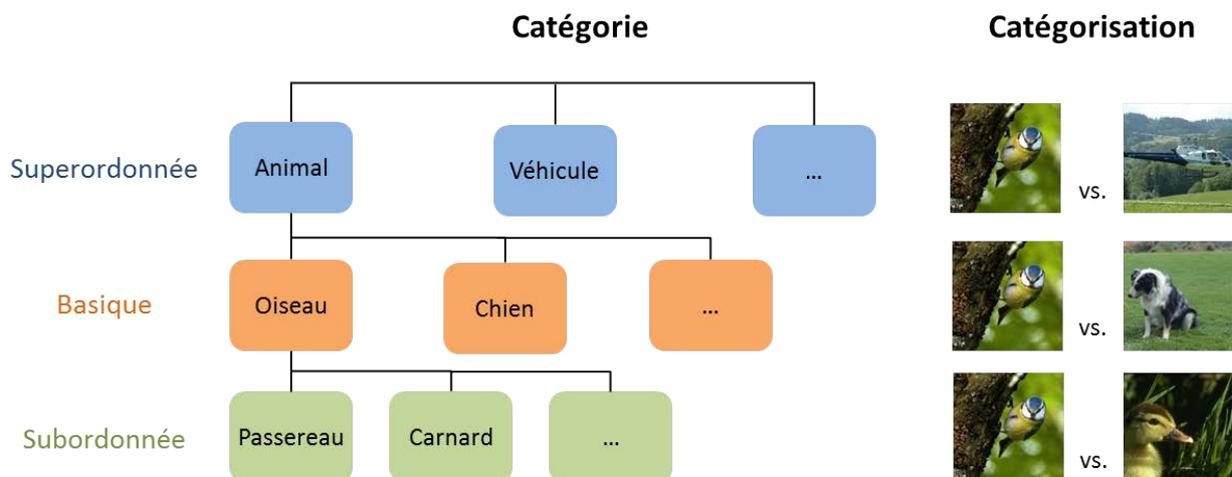


Figure 16. Exemples de différents niveaux de catégories et de tâches de catégorisation d'objets.

Il est important de noter tout de suite la différence entre une catégorisation oiseau vs. chien et une catégorisation oiseau vs. hélicoptère (exemples de la Figure 16). Ces deux types de tâches peuvent être considérés comme des catégorisations au niveau basique lorsque les participants doivent donner une réponse verbale (par exemple nommer des images en utilisant leur catégorie basique). Cependant, ce

n'est pas le cas lorsque les participants réalisent une tâche de catégorisation oui vs. non. Dans cette situation, ils doivent vérifier si l'image présentée correspond à la catégorie d'objet cible (qui peut être donnée à différents niveaux d'abstraction). Il est alors tout à fait possible de réaliser la tâche oiseau vs. voiture en effectuant une catégorisation superordonnée animal vs. véhicule puisque les participants n'ont pas besoin de savoir que c'est un oiseau ou une voiture pour répondre, il leur suffit de savoir que c'est un animal ou un véhicule. Au contraire, ils sont obligés de catégoriser les stimuli au niveau basique pour la tâche oiseau vs. chien (Bowers & Jones, 2008; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009).

En résumé, nous parlerons de trois niveaux de catégorisations : le niveau de base (catégoriser un oiseau par rapport à d'autres animaux), le niveau superordonné (catégoriser un animal parmi d'autres catégories superordonnées telles que véhicule), et le niveau subordonné (catégoriser un rapace parmi d'autres oiseaux).

2.1.1. La catégorie basique : point d'entrée de la reconnaissance visuelle

Le travail de Rosch et al., (1976) a été pionnier en ce qui concerne l'étude des différents niveaux de catégorisation. Leurs résultats suggèrent que les catégories basiques seraient les catégories d'objets les plus utilisées mais aussi les plus nécessaires pour le langage. En fait, les objets seraient catégorisés et nommés pour la première fois par les enfants au niveau basique. De plus, dans une de leur expérience, les auteurs ont demandé à des participants de lister le plus possible de caractéristiques s'appliquant à un objet particulier. Le nombre de caractéristiques listées était le plus élevé pour les catégories d'objets basiques. Selon les auteurs, la catégorie basique serait la plus informative : ses membres partageraient le plus d'attributs en communs et ces attributs leur permettraient aussi de les différencier des autres catégories basiques. Les catégories superordonnées et subordonnées sont peu informatives puisque dans le premier cas, seulement quelques caractéristiques sont partagées par leurs membres (elles sont trop inclusives) et dans le second cas, elles ne se distinguent pas bien des autres catégories subordonnées (beaucoup de leurs caractéristiques sont partagées avec d'autres catégories

subordonnées). En résumé, les catégories basiques seraient celles qui permettraient de coder le plus d'information au moindre coût cognitif.

Rosch et al., (1976) ont aussi testé dans leur étude la rapidité à laquelle les participants avaient accès aux catégories d'objets. Les participants entendaient le nom d'une catégorie d'objet et 500 ms plus tard une image était présentée sur un écran. Ils devaient vérifier si la catégorie de l'objet correspondait à celle de la catégorie entendue et donner leur réponse à l'aide d'un clavier d'ordinateur. Les résultats, reproduits quelques années plus tard par G. L. Murphy & Smith, (1982), montrent que les réponses étaient plus rapides pour les catégorisations au niveau basique (e.g. oiseau) que celles au niveau superordonné (e.g. animal) ou subordonné (e.g. mésange).

Le niveau à partir duquel les objets sont reconnus, a été appelé le « point d'entrée » de la reconnaissance par Jolicoeur, Gluck, & Kosslyn, (1984). Ce point d'entrée coïncide la plupart du temps avec la catégorie basique mais peut aussi se trouver au niveau de la catégorie subordonnée. C'est le cas par exemple des objets atypiques d'une catégorie basique : les participants catégorisent des images de pingouins plus rapidement en tant que pingouins c'est-à-dire leur catégorie subordonnée qu'en tant qu'oiseau, leur catégorie basique (Jolicoeur et al., 1984; G. L. Murphy & Brownell, 1985). Le point d'entrée de la reconnaissance est aussi influencé par le niveau d'expérience avec une catégorie spécifique. Rosch et al., (1976) avaient déjà remarqué que les experts dans un domaine, par exemple un mécanicien de maintenance aéronautique, donnaient beaucoup plus de caractéristiques s'appliquant aux avions que des novices (même si ces caractéristiques sont potentiellement disponibles pour les novices, elles sont le plus souvent ignorées). Dans une étude plus systématique, Tanaka & Taylor, (1991) ont montré que pour des experts en ornithologie, les catégories subordonnées devenaient aussi différenciées que les catégories basiques. Ce changement de point d'entrée était aussi évident dans les latences de réponses verbales qui devenaient aussi courtes que celles des catégories basiques.

De façon générale, la reconnaissance d'objets se ferait plus rapidement et avec une meilleure précision au niveau basique. Le traitement visuel des objets se ferait de façon hiérarchique : un objet serait d'abord catégorisé la majorité du temps au niveau basique (sauf dans les cas mentionnés ci-dessus d'un objet atypique ou d'une expertise dans un domaine particulier), alors que plus de temps serait nécessaire pour catégoriser le même objet à un niveau d'abstraction plus élevé (niveau superordonné) ou à un niveau plus détaillé (niveau subordonné).

2.1.2. La catégorisation basique n'est pas une étape nécessaire à la reconnaissance visuelle d'objets

L'hypothèse de la catégorie basique comme point d'entrée et donc comme étape nécessaire à la reconnaissance visuelle n'explique pas certains résultats neuropsychologiques. En effet, les patients atteints de démence sémantique perdent leurs connaissances des concepts intermédiaires ou basiques (comme le concept « chien ») tandis qu'ils gardent leurs connaissances de concepts plus généraux ou superordonnés (tel que « animal ») (Rogers & Patterson, 2007; Rogers, Ralph, Hodges, & Patterson, 2004; E K Warrington, 1975). La catégorie basique n'est donc pas une première étape nécessaire pour atteindre des niveaux de catégories superordonnées plus abstraits.

Rogers & Patterson, (2007) ont essayé d'expliquer cette contradiction en se basant sur la théorie PDP d'organisation des catégories qu'ils ont développée (voir partie II2 de l'introduction et Figure 17 panel gauche). D'après leur modèle, les représentations superordonnées devraient être activées avant les représentations plus spécifiques. Selon eux, l'avantage du niveau basique pourrait être renversé si les participants sont testés dans des conditions où le temps de réponse est contraint (Figure 17 panel droit).

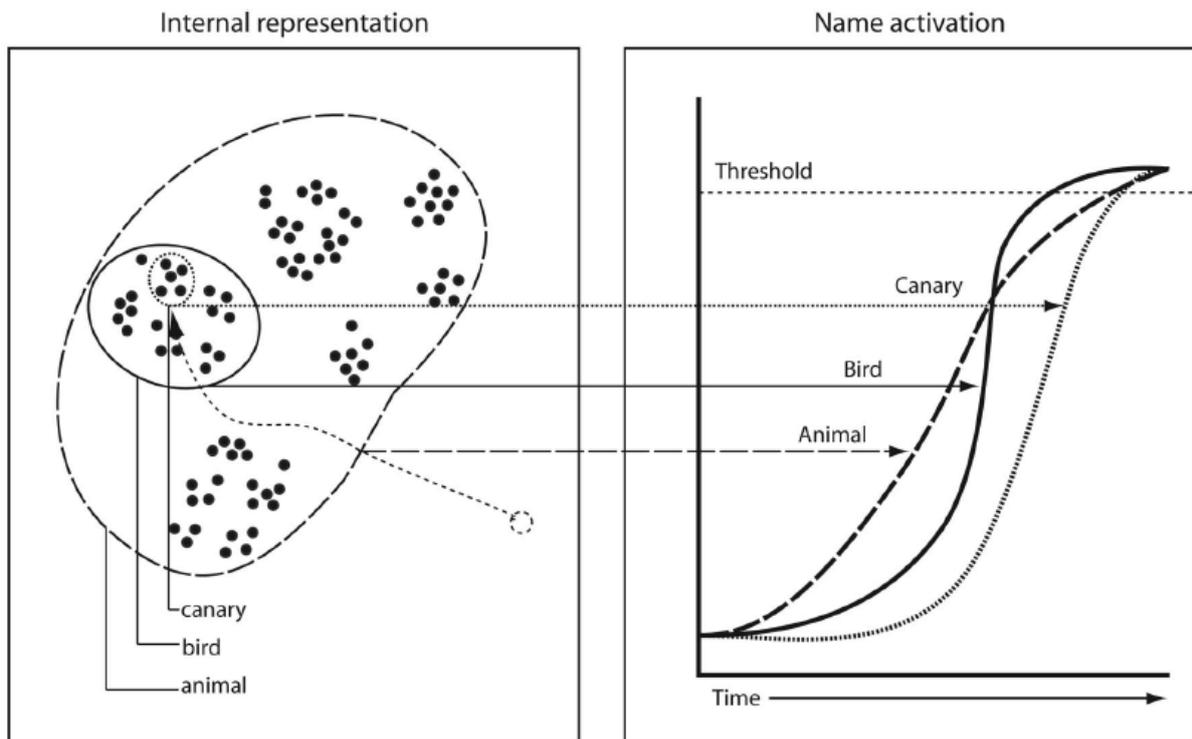


Figure 17. Représentation schématique montrant la prédiction de la théorie PDP de l'activation des noms au cours du temps pour différents niveaux de spécificité. Panel gauche : quand un stimulus visuel apparaît, l'état cognitif commence à partir d'un point neutre pour aller jusqu'à la représentation sémantique appropriée (par exemple un canari). Panel droit : pour atteindre le seuil de décision, le système passe d'abord par des régions de l'espace correspondant aux concepts superordonnés avant d'atteindre les concepts plus spécifiques. La courbe correspondant à la catégorie basique augmente plus rapidement (le nom est activé plus rapidement) puisqu'elle est la plus informative. Si le niveau de décision est plus bas (par exemple en utilisant un temps limite de réponse) alors on devrait voir apparaître un avantage de la catégorisation superordonnée par rapport à la catégorisation basique. (Rogers & Patterson, 2007)

Rogers & Patterson, (2007) ont testé leur hypothèse en utilisant le même type de procédure que Rosch et al., (1976) mais en ajoutant un temps de réponse limite. Les résultats montrent que pour des temps de réactions similaires, la précision des participants était meilleure pour les catégorisations superordonnées que pour les catégorisations basiques.

Depuis ces dernières années, les résultats de plusieurs études confortent cette hypothèse, en particulier ceux de Macé et al., (2009). En utilisant un paradigme de go/no-go, les auteurs de cette étude ont pu

montrer que les participants étaient plus rapides et plus précis pour catégoriser des objets visuels au niveau superordonné (e.g. animal) qu'au niveau basique (e.g. oiseau).

2.1.3. Présentation de l'article 1

Malgré l'augmentation du nombre d'articles en défaveur de l'hypothèse du point d'entrée de la reconnaissance au niveau basique, la possibilité d'un avantage des catégories superordonnées reste très discutée. Plusieurs critiques ont été émises dont une était l'utilisation d'un temps de présentation très court des stimuli dans l'expérience de Macé et al. (2009). En effet, les images étaient flashées à l'écran pendant seulement 20 ms. Ce temps de présentation bref pourrait favoriser le traitement des informations globales au détriment des informations plus précises et ainsi faire apparaître un avantage de la catégorisation superordonnée (Mack & Palmeri, 2011a).

Dans l'article 1, nous avons cherché à évaluer l'effet du temps de présentation des stimuli sur la rapidité d'accès aux différentes catégories d'objets visuels en présentant des stimuli pendant 20, 250 ou 500 ms. Nous avons utilisé des objets naturels (animaux) mais aussi artificiels (véhicules) afin de déterminer si nos résultats étaient généralisables à différents types de stimuli. Contrairement aux animaux, les véhicules ne sont apparus que très récemment dans notre environnement et il semblait donc intéressant de tester ces deux types de catégories d'objets. L'accès aux différents niveaux de catégorisation se fait avec la même hiérarchie temporelle quel que soit le temps de présentation. Cependant, les résultats de cette expérience soulèvent la question de l'effet de la diversité des stimuli au sein des catégories d'objets utilisées. Deux nouvelles expériences (expériences 2 et 3) ont donc été réalisées visant à tester l'influence de la diversité des catégories cibles et distracteurs. Nos résultats montrent que la diversité des stimuli semble bien jouer un rôle important dans les performances de catégorisation tandis que le temps de présentation et le mode de réponse (expérience 4) n'ont aucun effet.

2.2. Article 1: Stimulus duration and diversity do not reverse the advantage for superordinate representations: the animal is seen before the bird

Stimulus duration and diversity do not reverse the advantage for superordinate-level representations: the animal is seen before the bird

Marlène Poncet^{1,2} and Michèle Fabre-Thorpe^{1,2}

¹Centre de Recherche Cerveau et Cognition, UPS, Université de Toulouse, Toulouse, France

²CNRS CERCO UMR 5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France

Keywords: basic-level categorization, natural scenes, presentation time, stimulus diversity, visual categorization

Abstract

Basic-level categorization has long been thought to be the entry level for object representations. However, this view is now challenged. In particular, Macé *et al.* [M.J.-M. Macé *et al.* (2009) *PLoS One*, 4, e5927] showed that basic-level categorization (such as 'bird') requires a longer processing time than superordinate-level categorization (such as 'animal'). It has been argued that this result depends on the brief stimulus presentation times used in their study, which would degrade the visual information available. Here, we used a go/no-go paradigm to test whether the superordinate-level advantage could be observed with longer stimulus durations, and also investigated the impact of manipulating the target and distractor set heterogeneity. Our results clearly show that presentation time had no effect on categorization performance. Both target and distractor diversity influenced performance, but basic-level categories were never accessed faster or with higher accuracy than superordinate-level categories. These results argue in favor of coarse to fine visual processing to access perceptual representations.

Introduction

Since the late 1970s, our understanding of the temporal order in which humans access object categories has been dominated by the classic studies of Rosch *et al.* (1976) and others (Mervis & Rosch, 1981; Murphy & Smith, 1982). They reported that humans are faster at accessing object representations at the basic level (e.g. dog) than at the superordinate level (animal) or the subordinate level (e.g. poodle). This standard view implies a hierarchical processing scheme for visual stimuli, with the basic level as an entry step to reach a more abstract superordinate level of object recognition.

At odds with that theory, several studies have suggested that superordinate-level information might be available before the processing of more detailed information (Sugase *et al.*, 1999; Löw *et al.*, 2003; Large *et al.*, 2004; Martinovic *et al.*, 2008). Directly tackling the well-established basic-level entry point for object representation, Macé *et al.* (2009) used a go/no-go paradigm, and showed that participants were ~50 ms faster at categorizing a briefly flashed image as containing an animal than categorizing it as containing a bird or a dog. This superordinate-level advantage has been replicated in another study (Praž *et al.*, 2013), and also in scene gist categorization (Joubert *et al.*, 2007; Loschky & Larson, 2010; Kadar & Ben-Shahar, 2012).

In these experiments, images were masked or flashed very briefly. It has been argued that such fast presentations could emphasize coarse visual information over basic-level information (Mack & Palmeri, 2011). If the period of stimulus information uptake is interrupted, either because of masking or because of very short stimulus presentation, then perhaps only the features that can be extracted very rapidly can be used for further processing (Lamberts & Freeman, 1999; Lamberts, 2000). Another effect of brief presentation is that visual information gathered from the stimulus could be minimal and/or degraded (Pothos & Chater, 2002; Close & Pothos, 2012). Coarse representation might be relatively immune to such noise, whereas it could be deleterious for detailed representations, leading to a shift towards superordinate-level representations. With flashed stimuli, the superordinate-level advantage could thus be potentially a direct consequence of the restricted time available for the collection of information.

An alternative explanation might possibly account for the contradictory results for superordinate-level or basic-level advantages seen in different studies. It concerns the homogeneity vs. heterogeneity of the stimulus sets used as targets and distractors. Studies have shown that a 'dog' basic categorization is faster if the distractor category only includes exemplars from a different superordinate-level category (vehicle) than if the distractors include exemplars from the same animal superordinate-level category (Bowers & Jones, 2008; Macé *et al.*, 2009). Indeed, the similarity between and within categories might have a role in the speed of categorization (Mohan & Arun, 2012).

Correspondence: Michèle Fabre-Thorpe, ²CNRS, as above.
E-mail: michele.fabre-thorpe@cerco.ups-tlse

Received 12 November 2013, revised 30 December 2013, accepted 13 January 2014

By varying the presentation time of the stimuli, the type of object to be categorized, and the similarity within and between categories, we investigated the robustness of the superordinate-level advantage observed in go/no-go rapid visual categorization tasks.

Experiment 1: Stimulus duration

To test how stimulus duration would influence response latencies at different levels of visual categorization, we used an ultra-rapid categorization task introduced by Thorpe *et al.* (1996). Fast responses were encouraged by presenting a stimulus very briefly and by requiring go/no-go responses to be produced in a constrained time window (1 s). With this protocol, we investigated the effect of increasing the stimulus presentation time from 25 to 250 ms and 500 ms. If short stimulus duration emphasizes coarse processing, we should recover the basic-level advantage at longer presentation times. To generalize the results, both natural object categories (Experiment 1a) and artificial objects (Experiment 1b) were used.

Methods

Participants

Ten volunteers participated in Experiment 1a and Experiment 1b (four women, one left-handed; mean age, 29 years). Half of the participants performed Experiment 1a first, and the other half performed Experiment 1b first, with an interval of 1–5 days between the two experiments. All participants had normal or corrected to normal acuity, and provided written informed consent. The experiments received the approval of the French Ethical Committee (comité de protection des personnes Sud-Ouest et Outre-Mer I). They conformed to the Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the *British Medical Journal* (18 July 1964).

Apparatus and stimuli

Participants were seated in a dimly lit room, around 50 cm from a cathode ray tube screen (800 × 600 pixels; refresh rate, 120 Hz). Go response latencies were recorded via a response pad equipped with infrared diodes. The stimulus display and the response pad were synchronized with the refresh rate of the monitor.

All images used in the experiments were colored natural images chosen from the image sets used in previous studies (Macé *et al.*, 2009; Poncet *et al.*, 2012) and from the Internet. Objects were shown at various orientations, positions, and sizes (Fig. 1). The animal category included mammals, insects, birds, etc. The non-animal category included various vehicles and scenes without a specific foreground object (mountains, buildings, city, etc.). The content of each category is detailed in Table S1. Each participant saw the images (15° × 15° visual angle) only once. The level at which the image had to be categorized was randomly assigned, and the image status, target or distractor, depended on the level of categorization.

Procedure

Before the beginning of each 200-trial block, participants were told which category they had to respond to, and were trained on 20 images. To start a trial, subjects had to place their fingers on the response pad. A fixation square was then presented (600–1000 ms) at the center of a black screen, followed by the stimulus. Participants had to release their finger as quickly and as accurately as possible if the image contained a target object (they did not have to wait until the end of the stimulus presentation). Otherwise, they had to keep their finger on the response pad. The next trial started automatically after 1200 ms from stimulus onset. In each block, targets and distractors were equiprobable.

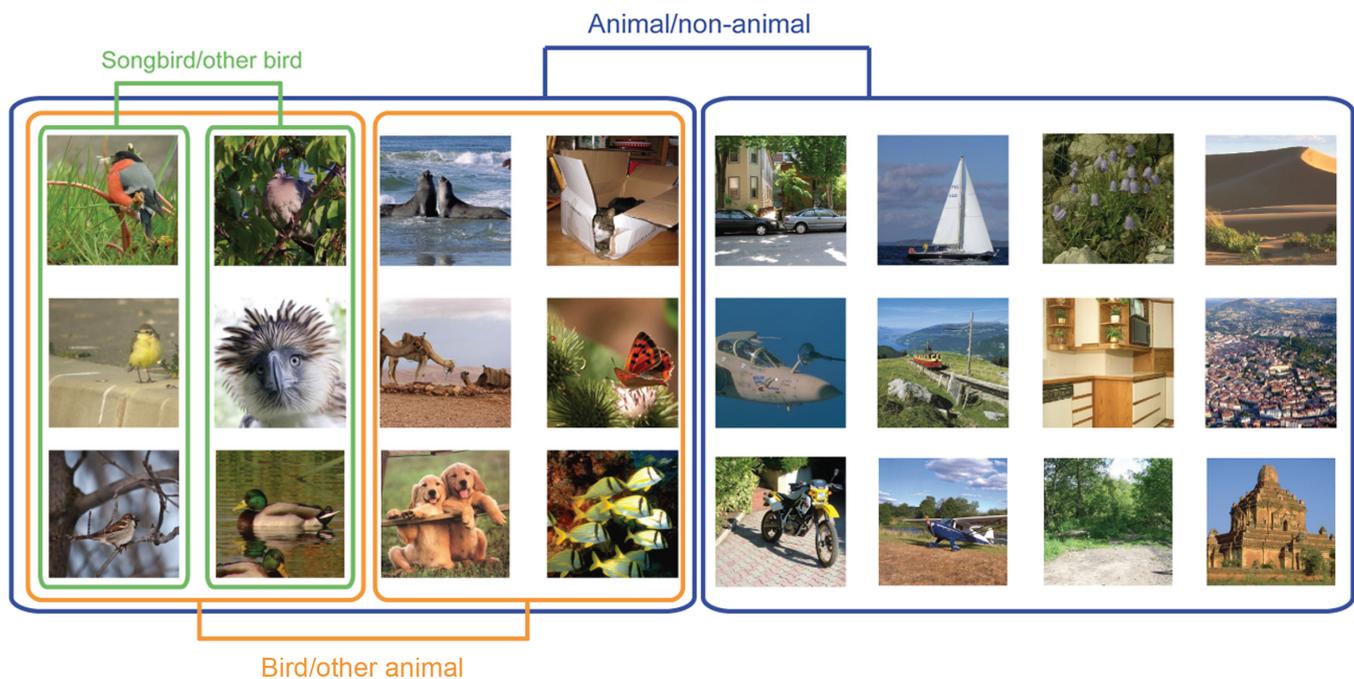


FIG. 1. Examples of images used in Experiment 1a. The three categorization tasks are framed according to the stimulus category: blue for the superordinate level; orange for the basic level; and green for the subordinate level. For each level of categorization, the target category is illustrated on the left and the distractor category on the right.

Experiment 1a. Participants had to categorize images of animals (vs. non-animals) at the superordinate level, birds (vs. other animals) at the basic level, and songbirds (vs. other birds) at the subordinate level. Each task was performed with three stimulus presentation times (25, 250, or 500 ms) for a total of nine blocks (three categorization levels \times three presentation times). Block order was randomly assigned to the participants.

Experiment 1b. The same procedure was used with artificial objects: vehicles (vs. non-vehicles) at the superordinate level, motorcycles (vs. other vehicles) at the basic level, and Harley motorcycles (vs. other motorcycles) at the subordinate level.

Data analysis

For each participant, behavior was assessed by the use of median RT for correct go responses and d' [corrected by Snodgrass & Corwin (1988)]. Anticipatory responses (RT < 150 ms) and RTs that were > 1 s were rejected from the analysis (0.13% of the trials). Repeated measures ANOVAs (level of categorization \times stimulus duration) were applied for each of the two experiments on median RT and d' . The same ANOVAs were performed without the subordinate-level results to enable a better comparison between superordinate-level and basic-level categorization performance. Sphericity of the data was assessed with Mauchly's sphericity test. Any violation was corrected with the Greenhouse–Geisser correction. When ANOVA results were significant for either the presentation time or the interaction factor, paired t -tests were applied to determine the relationship between the three presentation times. All t -test P -values reported in the article are those after Bonferroni correction. Results are given as average \pm standard error of the mean (SEM).

Results

Accuracy

Accuracy was very high in all categorization tasks: approximately 90 \pm 1% in Experiment 1a and 93 \pm 1% in Experiment 1b. In Experiment 1a, increasing presentation times increased the accuracy of performance at all levels of animal categorization ($P = 0.003$, no interaction) (see detailed results and statistical comparisons in Fig. 2A and Table 1). On the other hand, presentation time had no effect on the categorization of vehicles, with the exception of the subordinate level of motorcycles (Experiment 1b, Harley), in which accuracy was lower at 25 ms than at 250 ms ($t_9 = 3.3$, $P = 0.03$) and 500 ms ($t_9 = 5.7$, $P = 1.4 \times 10^{-4}$).

Relative to basic-level categorization, accuracy at the superordinate level was either higher for natural objects ($P = 0.006$) or similar for artificial objects. For all experiments, subordinate-level categorization was performed with a lower d' than for the other levels of categorization, but still with very high accuracy scores (> 80%).

RTs

The main aim of the study was to determine the impact of stimulus presentation time on the response latency at the three different levels of categorization. The results were very clear in all three experiments: stimulus presentation time had no effect on RTs (Table 1).

Comparison of response latencies at the three levels of categorization showed that subordinate-level categorization was always

performed most slowly ($P = 3.9 \times 10^{-11}$). Superordinate-level categorization was performed faster than basic-level categorization for animals as compared with birds (by 34 ± 3 ms; Experiment 1a; $P = 5.5 \times 10^{-7}$), but at similar speeds for vehicles as compared with motorcycles (Experiment 1b; 375 ± 14 ms and 373 ± 11 ms respectively; $P = 0.75$). As shown in Fig. 2B, the above results were valid for a large majority of the participants, and were not influenced by stimulus duration. The temporal advantage of superordinate-level over basic-level categorization for natural objects, and the absence of this for artificial objects, can be seen from the earliest responses. This is illustrated by the shift of RT distributions towards longer latencies for bird than for animal responses, and the overlap of RT distributions for vehicle and motorcycle responses (Fig. 3).

Discussion

Our results show no signs of a basic-level advantage that could appear with longer stimulus presentation times, which is at odds with many models of object recognition (Lamberts, 2000; Close & Pothos, 2012). Increasing presentation times increased accuracy for natural object categories, but response latencies remained unaffected.

One could argue that these results are the consequence of the restricted 1-s response time window that we used in our paradigm. It is very likely that, without time pressure, participants would slow down their responses to achieve higher accuracy for a 500-ms stimulus than for a 25-ms stimulus (accuracy trade-off). However, we show here that, with the same response criterion used in the different conditions, performance speed was not affected by the duration of the stimulus.

We found a superordinate-level advantage for animal over bird categorization but not for vehicle over motorcycle categorization. On the one hand, animals might be considered to constitute a very special category of biologically pertinent objects that rely on faster hard-wired neural mechanisms, possibly tuned by ancestral priorities (New *et al.*, 2007). On the other hand, animal and vehicle stimuli were categorized at the superordinate level with similar accuracies and RTs (RT, paired t -test $t_9 = 0.04$, $P = 0.9$; d' , $t_9 = 0.97$, $P = 0.35$), a result that replicates the findings of a study by VanRullen & Thorpe (2001). It might, then, be possible that categorization performance for motorcycles reached the level of a superordinate-level categorization because this basic vehicle category is very homogeneous as compared with the bird category.

To test whether target heterogeneity could play a role in the results obtained, in Experiment 2 we used a more diverse basic category of vehicles (car). If categorization performance were similar for vehicles and cars, such as in Experiment 1b with vehicles and motorcycles, this would suggest special visual processing of animals. In contrast, if a superordinate-level advantage were observed, such as in Experiment 1a with animals and birds, this would indicate a role of stimulus diversity in categorization performance.

Experiment 2: Target category diversity

Methods

Experiment 2 was performed by 13 new volunteers (seven women, one left-handed; mean age, 25 years). The protocol was similar to that of Experiment 1, with cars (vs. other vehicles) at the basic level

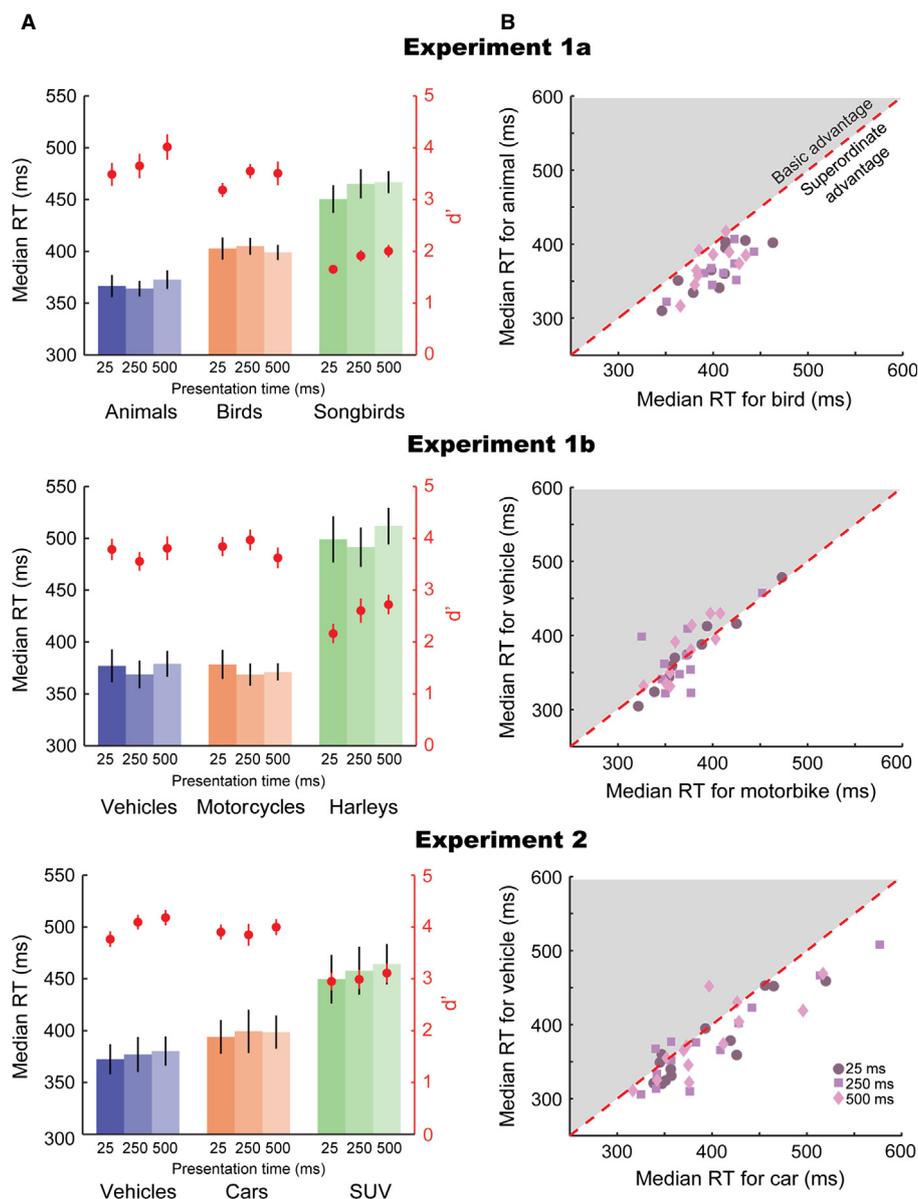


FIG. 2. Results of Experiments 1 and 2 with animal (Experiment 1a) or vehicle (Experiments 1b and 2) categories. (A) Bars illustrate the median RT at each presentation time for superordinate-level (blue), basic-level (orange) and subordinate-level (green) categorization. Red circles denote accuracy (d' , see right y-axis). Error bars represent the SEM. (B) The median RT in the basic-level categorization is compared with the median RT in the superordinate-level categorization. Each symbol represents the performance of an individual participant for presentation times of 25 ms (dark circles), 250 ms (squares), and 500 ms (light diamonds). Symbols falling in the shaded area indicate a basic-level advantage.

and sport utility vehicles (SUVs) (vs. other cars) at the subordinate level.

Results

As observed in Experiment 1, accuracy was very high in all categorization tasks ($94 \pm 3\%$). Participants' accuracy (d') was similar for all presentation times (Fig. 2; Table 1), but lower for the subordinate-level categorization (SUV) than for other levels of categorization (vehicle and car). Thus, the pattern of d' results was very similar to that found in Experiment 1b, but that of RT results was not. SUVs were categorized more slowly than cars and vehicles, but, more importantly, cars were also categorized more slowly than

vehicles (by 21 ± 6 ms; $P = 0.004$). Thus, we found a superordinate-level advantage with vehicles/cars as categories of artificial object. This effect was not affected by presentation time ($P = 0.65$), and was found in most of the participants (Fig. 2B).

Discussion

In this experiment, we tested whether we could replicate the results obtained in Experiment 1b when using a less homogeneous basic category of vehicle. As in Experiment 1b, our results showed no effect of presentation time. However, whereas vehicles and motorcycles were categorized with similar speed and accuracy in Experiment 1b, vehicles were categorized faster than cars in Experiment 2. This

TABLE 1. Statistical results of experiments 1 and 2. Repeated measures ANOVA between different levels of categorization were applied to median RT and d'

	Effect	RT			d'		
		F	P	η^2	F	P	η^2
Superordinate-level vs. basic-level categorization							
Experiment 1a: animal, bird	Category	$F_{1,9} = 155.69$	5.5×10^{-7}	0.94	$F_{1,9} = 12.60$	0.006	0.58
	Presentation time	$F_{2,18} = 0.06$	0.94	0.01	$F_{2,18} = 5.11$	0.02	0.36
	Interaction	$F_{2,18} = 1.25$	0.31	0.12	$F_{2,18} = 1.03$	0.38	0.10
Experiment 1b: vehicle, motorcycle	Category	$F_{1,9} = 0.11$	0.75	0.01	$F_{1,9} = 1.85$	0.21	0.17
	Presentation time	$F_{2,18} = 0.87$	0.44	0.09	$F_{2,18} = 0.32$	0.73	0.03
	Interaction	$F_{2,18} = 0.87$	0.43	0.09	$F_{2,18} = 3.21$	0.06	0.26
Experiment 2: vehicle, car	Category	$F_{1,12} = 12.52$	0.004	0.51	$F_{1,12} = 1.83$	0.20	0.13
	Presentation time	$F_{2,24} = 0.44$	0.65	0.04	$F_{1,17} = 2.08$	0.16	0.15
	Interaction	$F_{2,24} = 0.10$	0.90	0.01	$F_{2,24} = 1.88$	0.17	0.13
Superordinate-level vs. basic-level vs. subordinate-level categorization							
Experiment 1a: animal, bird, songbird	Category	$F_{1,17} = 120.00$	1.6×10^{-7}	0.93	$F_{1,12} = 137.59$	3.0×10^{-8}	0.94
	Presentation time	$F_{2,18} = 0.99$	0.39	0.10	$F_{2,18} = 8.32$	0.003	0.48
	Interaction	$F_{4,36} = 1.57$	0.20	0.15	$F_{4,36} = 0.78$	0.54	0.08
Experiment 1b: vehicle, motorcycle, Harley	Category	$F_{2,18} = 130.64c$	1.9×10^{-11}	0.94	$F_{2,18} = 77.29$	1.5×10^{-9}	0.90
	Presentation time	$F_{2,18} = 1.37$	0.29	0.13	$F_{2,18} = 1.32$	0.29	0.13
	Interaction	$F_{4,36} = 1.71$	0.17	0.16	$F_{4,36} = 4.41$	0.005	0.33
Experiment 2: vehicle, car, SUV	Category	$F_{1,11} = 35.28$	4.3×10^{-6}	0.75	$F_{2,44} = 40.21$	2.2×10^{-8}	0.77
	Presentation time	$F_{2,24} = 1.34$	0.28	0.10	$F_{2,44} = 2.83$	0.08	0.19
	Interaction	$F_{4,48} = 0.28$	0.89	0.02	$F_{4,48} = 1.21$	0.32	0.09

Significant values are in bold.

superordinate-level advantage replicates the results observed with animal and bird categories in Experiment 1a, albeit with a weaker effect. As already suggested, the difference between motorcycle and car categorization performance could be explained by the diversity of exemplars in the target category (cars are more diverse than motorcycles). Furthermore, the perceptual distance between targets and distractors might also have an effect on processing time (Mohan & Arun, 2012). The stronger superordinate-level advantage observed in Experiment 1a could thus result from the very large stimulus diversity in the animal kingdom and in each animal basic category. In contrast, the vehicle category is less diverse, and basic categories such as car and motorcycle are easily distinguishable from other vehicles. In Experiment 3, we further tested the influence of increasing and decreasing exemplar similarity on basic-level categorization performance. To this end, we manipulated the homogeneity of the animal image set, as this category is more diverse than the vehicle category.

Experiment 3: Image set homogeneity

Methods

Fourteen participants (11 women, one left-handed; mean age, 25 years) categorized images in a go/no-go paradigm similar to the one used in Experiment 1, but with stimuli always displayed for 250 ms. The stimulus diversity in the target and distractor sets was manipulated for basic-level categorization while instructions were kept the same (release your finger as quickly and as accurately as possible when you see a bird). The target category 'bird' could be restricted to birds of prey or could include different bird species. The distractor category 'non-bird animal' could be restricted to dogs or include any kind of animal. Therefore, basic-level categorization could be performed in three different conditions. Target and distractor categories could both be restricted (condition RR: bird of prey vs. dog), the target category could be broadly open with a restricted distractor category (condition BR: bird vs. dog), or both target and

distractor categories could be broadly opened (condition BB: bird vs. non-bird animal). Participants were not told about the differences in the experimental design. They were also tested at the superordinate level (animal) and subordinate level (with bird of prey as target). The distractor category was restricted to vehicle at the superordinate level, and to songbird at the subordinate level. Participants performed a total of five blocks of 200 trials (three blocks at the basic level, RR-BR-BB; one block at the superordinate level; and one block at the subordinate level).

The median RT for correct go responses and d' were calculated for each participant. Paired t -test comparisons between the different types of categorization were applied (four comparisons in total: superordinate vs. BB condition, BB vs. BR condition, BR vs. RR condition, and RR vs. subordinate condition). All t -test P -values reported are those after Bonferroni correction. In order to determine the effect of restricting the distractor set at each level of categorization, we conducted an ANOVA on the results of Experiment 1a (when stimuli were presented for 250 ms) and Experiment 3 (with the BR condition at the basic level), using the level of categorization as a within-subject factor and the group of participants as a between-subject factor.

Results

The central manipulation of Experiment 3 was modulation of the homogeneity of the target and distractor sets when subjects were asked to perform the same basic-level categorization task. Increasing the diversity of the target set while keeping the distractor set restricted (comparison RR vs. BR) induced a significant 21-ms RT increase ($t_{13} = 3.3$, $P = 0.02$), but had no effect on d' ($t_{13} = 1.7$, $P = 0.46$) (Fig. 4). A further increase in stimulus diversity in the distractor set (comparison BR vs. BB) decreased d' by 0.7 ($t_{13} = 3.9$, $P = 0.007$), and had a tendency to increase RT ($t_{13} = 2.6$, $P = 0.09$).

Our results showed the same temporal dynamics for accessing the three levels of object representation as in Experiment 1. The

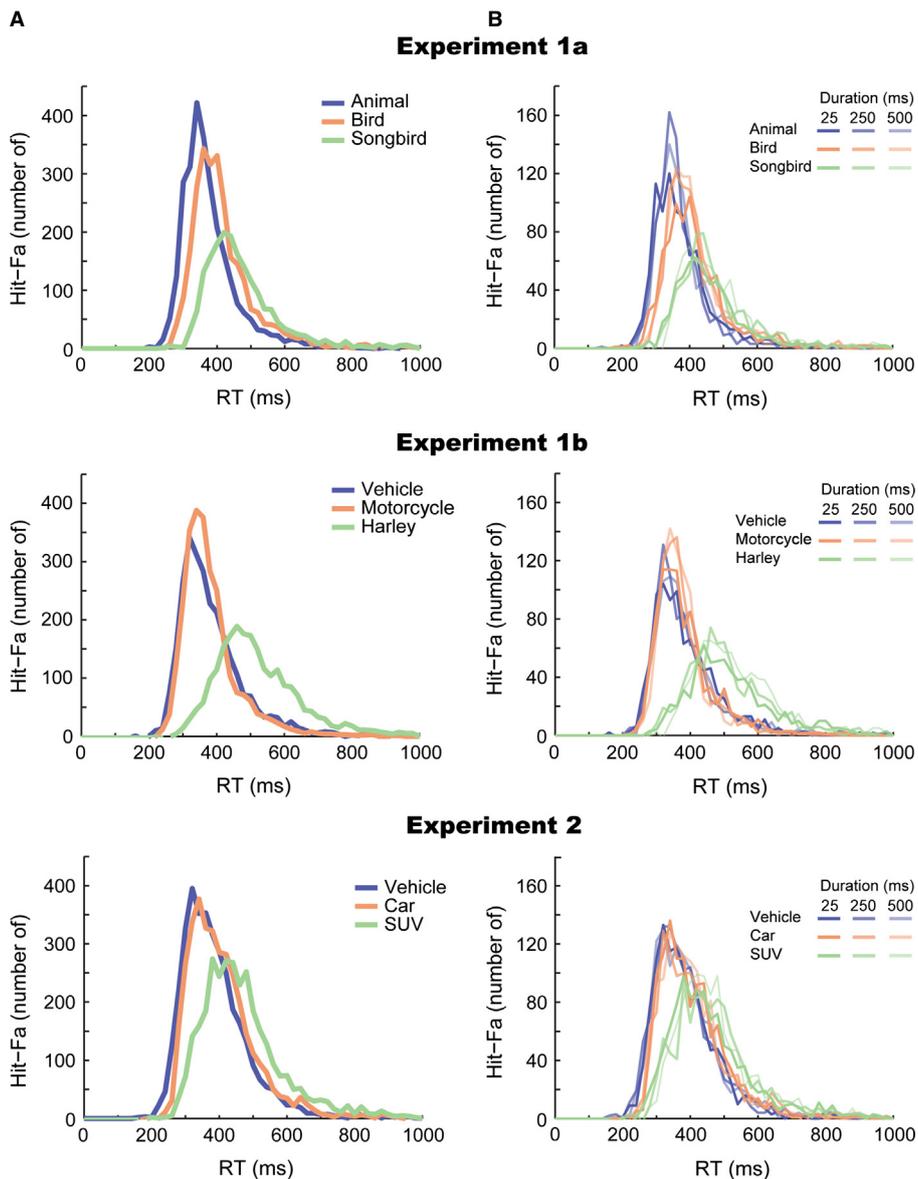


FIG. 3. RT distribution in Experiments 1a, 1b and 2 for superordinate-level (blue), basic-level (orange) and subordinate-level (green) categorization. Each RT distribution was obtained by subtracting the number of false alarms (Fa) from the number of Hits (Hit) in a 20-ms bin. The RT distribution is shown for three presentation times separately (B) or pooled (A). The longer the presentation time of the stimulus, the lighter the color of the curve.

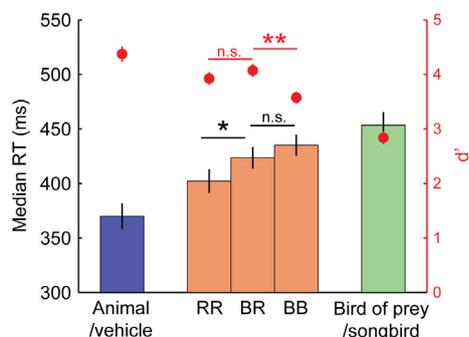


FIG. 4. RT (bars) and d' (red circles) in Experiment 3 for superordinate-level (blue), basic-level (orange) and subordinate-level (green) categorization. The stimulus diversity at the basic level was either: restricted for the target and distractor sets (RR); broad for the target and restricted for the distractor set (BR); or broad for the target and distractor sets (BB). Error bars represent the SEM. n.s., not significant. * $P < 0.05$, ** $P < 0.01$.

superordinate-level advantage was still present even when basic-level categorization was made easier by restricting the diversity of the targets and distractors (comparison with RR condition). Participants were 32 ± 8 ms faster ($t_{13} = 3.78$, $P = 0.009$) and more accurate ($t_{13} = 3.35$, $P = 0.02$) in categorizing animals vs. vehicles than in categorizing birds of prey vs. dogs. In accordance with the results of Experiment 1, subordinate-level categorization was always performed less accurately than basic-level categorization. Participants performed bird of prey vs. songbird categorization with similar RTs ($t_{13} = 2.19$, $P = 0.19$) but with much lower d' values ($t_{13} = 8.86$, $P = 2.9 \times 10^{-6}$) than bird vs. animal categorization (BB condition).

An interesting comparison can be performed between the results obtained in Experiment 1a at 250 ms and Experiment 3. In Experiment 1a, whatever the level of categorization, the target and distractor sets were always broadly open. On the other hand, in Experiment 3, performance was investigated at all levels in the BR condition (distractors were restricted to vehicles at the superordinate

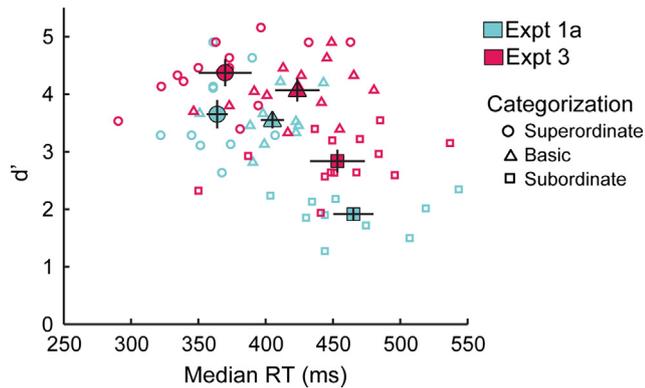


FIG. 5. Individual median RT as a function of d' in Experiment 1a (blue) and Experiment 3 (red) for categorizations at the superordinate level (circles), basic level (triangles), and subordinate level (squares). Filled shapes represent the average of performance in each condition. The distractor set diversity was broad in Experiment 1a but restricted in Experiment 3.

level and to songbirds at the subordinate level). Therefore, we could compare Experiment 1a and Experiment 3 to investigate the effect of restricting the distractor set at the three levels of categorization (Fig. 5). We found an effect of the level of categorization for both RT ($F_{2,44} = 109.2$, $P = 8.7 \times 10^{-18}$, $\eta^2 = 0.85$) and d' ($F_{2,44} = 123.3$, $P = 9.2 \times 10^{-19}$, $\eta^2 = 0.85$). These results were expected because, in both experiments, we found a superordinate-level advantage (therefore, no further *post hoc* analyses were performed). As observed at the basic level, increasing the diversity in the distractor set had no effect on RT ($F_{1,22} = 0.1$, $P = 0.77$; no interaction, $F_{2,44} = 3.0$, $P = 0.06$). However, d' values were significantly higher in Experiment 2 than in Experiment 1a ($F_{1,22} = 25.7$, $P = 4.4 \times 10^{-5}$, $\eta^2 = 0.54$) at all levels of categorization (no interaction, $F_{2,44} = 0.9$, $P = 0.39$).

Discussion

In this experiment, we again found that categorization at the superordinate level was performed faster and more accurately than categorization at the basic level or subordinate level. Increasing heterogeneity in the target or distractor sets had different behavioral effects: large target diversity increased RT, whereas large distractor diversity decreased d' .

In our results, the speed of categorization was mostly driven by modulation of the target set but not by modulation of the distractor set. One could argue that this is attributable to the paradigm used. In a go/no-go procedure, participants are more focused on the target category than on the distractor category, whereas in a yes/no paradigm, RTs could be adjusted for both target and distractor sets, making it more likely that a basic-level advantage could emerge. We tested this hypothesis in Experiment 4.

Experiment 4: Type of response

Methods

Nine participants (two women, one left-handed; mean age, 26 years) performed Experiment 4 in a procedure similar to that used in Experiment 1. Participants were tested in a go/no-go task and in a yes/no task on superordinate-level (animal vs. non-animal) and basic-level (bird vs. non-bird animal) categorization. Stimuli were

always displayed for 250 ms (as in Experiment 3), and in both response conditions participants reported their responses on a keyboard. They had to keep their finger on the space bar and release it in the go/no-go condition, whereas in the yes/no condition they had to press the 'c' or the 'n' key for a 'yes' or a 'no' response, respectively. Median RTs for correct go and yes responses, as well as d' , were calculated for each participant. Repeated measures ANOVAS (level of categorization \times type of response) were then applied.

Results

Animal categorization was performed faster ($F_{1,8} = 39.5$, $P = 2.4 \times 10^{-4}$, $\eta^2 = 0.83$) than bird categorization, but with similar d' ($F_{1,8} = 1.56$, $P = 0.25$) (Fig. 6). Performance (RT and d') was lower with yes/no responses than with go/no-go responses (RT, $F_{1,8} = 25.5$, $P = 0.001$, $\eta^2 = 0.76$; d' , $F_{1,8} = 23.7$, $P = 2.3 \times 10^{-4}$, $\eta^2 = 0.83$), as found previously (Bacon-Macé *et al.*, 2007), but this difference was the same for superordinate-level and basic-level categorization (no interaction: RT, $F_{1,8} = 1.4 \times 10^{-6}$, $P = 1.00$; d' , $F_{1,8} = 0.01$, $P = 0.93$), and was present from the earliest responses (Fig. S1).

Discussion

We found a 32 ± 6 -ms superordinate-level advantage for both go and yes correct responses. Therefore, the adjustment of response thresholds for the target category does not depend on a specific type of response.

General discussion

The first goal of the study was to investigate alternative explanations that have been put forward to explain the superordinate-level advantage found in fast visual categorization tasks. According to Mack & Palmeri (2011), 'with limited exposure, it is likely that only relatively coarse and potentially salient visual properties of an image are encoded ... additional exposure would be necessary to encode more detailed features required for fast and accurate basic- or subordinate-level categorization'. Thus, superordinate-level representations would be favored because of the brief presentation of stimuli, which imposes temporal constraints on the collection of visual information. Our results showed that increasing stimulus duration did not prevent the superordinate-level advantage. Superordinate-level categories were always accessed with shorter processing times than other levels of category (except for the motorcycle category; but see late).

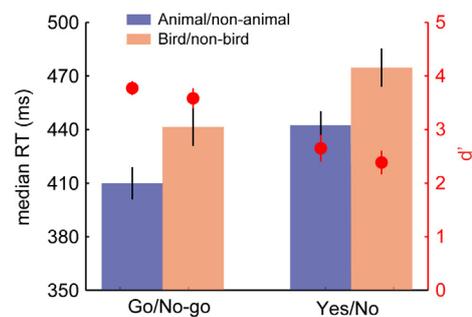


FIG. 6. Results of Experiment 4. Participants performed superordinate-level (blue) and basic-level (orange) categorizations in either a go/no-go (left side) or a yes/no (right side) task. Bars represent median RT and red circles d' . Error bars represent the SEM.

Furthermore, the superordinate-level advantage was not reversed by manipulating the diversity of exemplars in the target and the distractor sets (Experiment 3) or by using a new response mode (Experiment 4, yes/no). Thus, our results show that the superordinate-level advantage seen in rapid visual categorization tasks is robust, and is not a result of the coarser information quality associated with briefly flashed stimuli; the animal is seen earlier than the bird, whatever the quality of the gathered information.

The results of Experiment 3 showed that, whereas higher target heterogeneity induced longer RTs, higher heterogeneity among distractors decreased accuracy. When the target set is restricted, the number of relevant features may be smaller and more reliable than with a large diversity of target exemplars. One possibility is that the areas engaged in category decisions change their read-out of the visual cortical responses, and only focus on these relevant features (Li *et al.*, 2009). A non-exclusive alternative is that object-based attention could modulate bottom-up visual processing (Reynolds & Heeger, 2009; Miller *et al.*, 2011). The neuronal threshold could be adjusted to pertinent features, depending on the goal (look for the bird) and the context (target and distractor heterogeneity) of the task. After a correct trial, the amount of information required to reach the decision threshold will be reduced, until it increases again when an incorrect response is produced. These fluctuations will quickly stabilize, and the resulting decision threshold could be lower for a homogeneous set than for a heterogeneous set. This hypothesis might also explain why RTs for motorcycles and vehicles were similar in Experiment 1b. The shorter RTs for motorcycles than for other basic-level object categories might be explained by higher homogeneity in the motorcycle set of images: different motorcycles are perceptually more similar than different cars. Thus, at a given level of categorization (e.g. basic), increased homogeneity in the target set (while the same distractor set is kept) could correlate with shorter RTs. These assumptions could be tested in a computational model.

On the other hand, it seems that accuracy scores depend mostly on the number of features shared between targets and distractors. When the distractor category is large, the probability of finding a relevant feature in a distractor image increases, which could lead to more false alarms. This result fits with the study of Macé *et al.* (2009), where they found that false alarms were mostly seen with distractors that were similar to the target category. When participants were performing a dog categorization task, false alarms often occurred for wolves and foxes, whereas when participants were performing a bird categorization task, false alarms occurred for fish and insects.

It should be noted that the stimuli were unmasked in the present study, as the goal was to reproduce the experimental conditions used in most ultra-rapid categorization studies. Indeed, a mask would have interrupted, disturbed and/or limited the processing of the stimulus, when our goal was to allow maximal processing of the information collected with longer stimulus presentation time (using situations close to normal viewing conditions). Although we cannot draw any conclusions about whether the superordinate-level categorization is a necessary stage in reaching basic-level representations (Mack *et al.*, 2009), we clearly show that superordinate-level representations can be accessed faster than basic-level ones, which is at odds with the basic-level advantage often reported.

The go/no-go paradigm used in the present study appears to be a good protocol with which to study the time course of perceptual categorization. In contrast, the verification tasks typically used in experiments where a basic-level advantage has been reported (e.g. Rosch *et al.*, 1976) might not reflect the steps of pure visual processing. For

each trial, the name of the target category was presented before the stimulus, meaning that the task also involves reading and language areas. The instruction probably triggers iterative loops between the different cortical areas, in order to prepare and modulate the threshold of the neurons involved in the task, which could favor basic-level over superordinate-level responses. However, in a go/no-go paradigm, the instruction is only given at the beginning of the block. Object processing might be independent from language areas, as the neuronal threshold in the visual system will be adjusted following correct and incorrect responses. Indeed, RTs in ultra-rapid categorization tasks are much faster than in any verification task, and these ultra-rapid categorization tasks can be performed by monkeys that do not have access to language and have performance levels very similar to those of humans (Macé *et al.*, 2010; Fize *et al.*, 2011). Whatever the level of categorization, the system is in an optimal state to process upcoming visual information, and our results clearly show faster access to superordinate-level representations. Moreover, they also show that this superordinate-level advantage does not result from degraded information that would favor coarse representations, as it is very robust to increasing stimulus duration.

Another explanation for the superordinate-level advantage comes from electrophysiological and functional magnetic resonance imaging results. It has been shown that the brain activity elicited in the infero-temporal cortex by a visual object is most dissimilar between animate and inanimate objects (Kiani *et al.*, 2007; Kriegeskorte *et al.*, 2008). Superordinate-level categorization tasks, which compare animate and inanimate object categories, might directly rely on this brain organization, and would thus be performed faster than basic-level categorization tasks, which compare different animate exemplars.

To conclude, this study shows that the advantage of the superordinate level of categorization over basic-level representations revealed by rapid categorization tasks persists even when long stimulus durations up to 500 ms are used. It also reveals that increasing exemplar diversity in the target and distractor exemplar sets does not reverse the superordinate-level advantage but has differential effects on performance. Our results are in favor of the idea that visual processing first accesses coarse representations, before allowing finer differentiation. This might be supported by the clear difference between the brain activities elicited by animate and inanimate objects.

Supporting Information

Additional supporting information can be found in the online version of this article:

Table S1. Description of the content of each target and distractor images set in Experiment 1a, Experiment 1b and Experiment 2 for the three categorization tasks.

Fig. S1. RT distributions (20-ms bins) in Experiment 4 for superordinate-level (blue) and basic-level (red) categorization in either a go/no-go task (darker curves) or a yes/no task (lighter curves).

Acknowledgements

We thank Simon Thorpe for helpful comments on the manuscript. The authors declare no conflict of interests.

Abbreviations

RT, reaction time; SEM, standard error of the mean; SUV, sport utility vehicle.

References

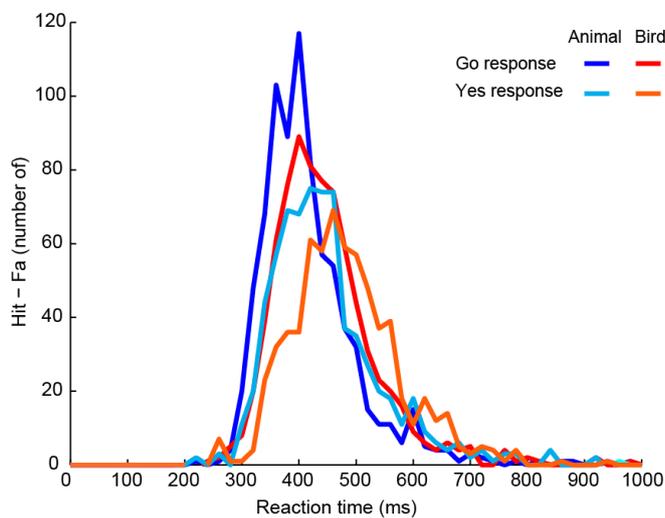
- Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M. & Thorpe, S.J. (2007) Effects of task requirements on rapid natural scene processing: from common sensory encoding to distinct decisional mechanisms. *J. Exp. Psychol. Human.*, **33**, 1013–1026.
- Bowers, J.S. & Jones, K.W. (2008) Detecting objects is easier than categorizing them. *Q. J. Exp. Psychol.*, **61**, 552–557.
- Close, J. & Pothos, E.M. (2012) 'Object categorization: reversals and explanations of the basic-level advantage' (Rogers & Patterson, 2007): a simplicity account. *Q. J. Exp. Psychol.*, **65**, 1615–1632.
- Fize, D., Cauchoix, M. & Fabre-Thorpe, M. (2011) Humans and monkeys share visual representations. *Proc. Natl. Acad. Sci. USA*, **108**, 7635–7640.
- Joubert, O.R., Rousselet, G.A., Fize, D. & Fabre-Thorpe, M. (2007) Processing scene context: fast categorization and object interference. *Vision Res.*, **47**, 3286–3297.
- Kadar, I. & Ben-Shahar, O. (2012) A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *J. Vision*, **12**, 16.
- Kiani, R., Esteky, H., Mirpour, K. & Tanaka, K. (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J. Neurophysiol.*, **97**, 4296–4309.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K. & Bandettini, P.A. (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, **60**, 1126–1141.
- Lamberts, K. (2000) Information-accumulation theory of speeded categorization. *Psychol. Rev.*, **107**, 227–260.
- Lamberts, K. & Freeman, R.P.J. (1999) Categorization of briefly presented objects. *Psychol. Res.*, **62**, 107–117.
- Large, M.-E., Kiss, I. & McMullen, P.A. (2004) Electrophysiological correlates of object categorization: back to basics. *Brain Res. Cogn. Brain Res.*, **20**, 415–426.
- Li, S., Mayhew, S.D. & Kourtzi, Z. (2009) Learning shapes the representation of behavioral choice in the human brain. *Neuron*, **62**, 441–452.
- Loschky, L.C. & Larson, A.M. (2010) The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Vis. Cogn.*, **18**, 513–536.
- Löw, A., Bentin, S., Rockstroh, B., Silberman, Y., Gomolla, A., Cohen, R. & Elbert, T. (2003) Semantic categorization in the human brain spatiotemporal dynamics revealed by magnetoencephalography. *Psychol. Sci.*, **14**, 367–372.
- Macé, M.J.-M., Joubert, O.R., Nespoulous, J.-L. & Fabre-Thorpe, M. (2009) The time-course of visual categorizations: you spot the animal faster than the bird. *PLoS ONE*, **4**, e5927.
- Macé, M.J.-M., Delorme, A., Richard, G. & Fabre-Thorpe, M. (2010) Spotting animals in natural scenes: efficiency of humans and monkeys at very low contrasts. *Anim. Cogn.*, **13**, 405–418.
- Mack, M.L. & Palmeri, T.J. (2011) The timing of visual object categorization. *Front. Psychol.*, **2**, 165.
- Mack, M.L., Wong, A.C.-N., Gauthier, I., Tanaka, J.W. & Palmeri, T.J. (2009) Time course of visual object categorization: fastest does not necessarily mean first. *Vision Res.*, **49**, 1961–1968.
- Martinovic, J., Gruber, T. & Müller, M.M. (2008) Coding of visual object features and feature conjunctions in the human brain. *PLoS ONE*, **3**, e3781.
- Mervis, C.B. & Rosch, E. (1981) Categorization of natural objects. *Annu. Rev. Psychol.*, **32**, 89–115.
- Miller, B.T., Vytlačil, J., Fegen, D., Pradhan, S. & D'Esposito, M. (2011) The prefrontal cortex modulates category selectivity in human extrastriate cortex. *J. Cognitive Neurosci.*, **23**, 1–10.
- Mohan, K. & Arun, S.P. (2012) Similarity relations in visual search predict rapid visual categorization. *J. Vision*, **12**, 19.
- Murphy, G.L. & Smith, E.E. (1982) Basic-level superiority in picture categorization. *J. Verb. Learn. Verb. Be.*, **21**, 1–20.
- New, J., Cosmides, L. & Tooby, J. (2007) Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc. Natl. Acad. Sci. USA*, **104**, 16598–16603.
- Poncet, M., Reddy, L. & Fabre-Thorpe, M. (2012) A need for more information uptake but not focused attention to access basic-level representations. *J. Vision*, **12**, 15.
- Pothos, E.M. & Chater, N. (2002) A simplicity principle in unsupervised human categorization. *Cognitive Sci.*, **26**, 303–343.
- Praß, M., Grimsen, C., König, M. & Fahle, M. (2013) Ultra rapid object categorization: effects of level, animacy and context. *PLoS ONE*, **8**, e68051.
- Reynolds, J.H. & Heeger, D.J. (2009) The normalization model of attention. *Neuron*, **61**, 168–185.
- Rosch, E., Mervis, C., Gray, W., Johnson, D. & Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychol.*, **8**, 382–439.
- Snodgrass, J.G. & Corwin, J. (1988) Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.*, **117**, 34–50.
- Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, **400**, 869–873.
- Thorpe, S.J., Fize, D. & Marlot, C. (1996) Speed of processing in the human visual system. *Nature*, **381**, 520–522.
- VanRullen, R. & Thorpe, S.J. (2001) Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, **30**, 655–668.

SUPPORTING INFORMATION

		Superordinate	Basic	Subordinate
Exp1a	Target	25% non-bird animals 12,5% non-songbird birds 12,5% songbirds	25% non-songbird birds 25% songbirds	50% songbirds
	Distractor	25% neutral 25% vehicles	50% non-bird animals	50% non-songbird birds
Exp1b	Target	25% non-motorbike vehicles 12,5% non-cruiser motorbikes 12,5% cruisers	25% non-cruiser motorbikes 25% cruisers	50% cruisers
	Distractor	25% neutral 25% animals	50% non-motorbike vehicles	50% non-cruiser motorbikes
Exp2	Target	25% non-car vehicle 12,5% non-SUV cars 12,5% SUVs	25% non-SUV cars 25% SUVs	50% SUVs
	Distractor	25% neutral 25% animals	50% non-car vehicles	50% non-SUV cars

Supplementary Table1: Description of the content of each target and distractor images set in

Exp1a, Exp1b and Exp2 for the 3 categorization tasks.



Supplementary Figure 1: RT distributions (20ms bins) in Exp4 for superordinate (blue) and basic (red) level of categorization either in a go/no-go task (darker curves) or in a yes/no task (lighter curves).

Presentation time does not affect the superordinate-level advantage in ultra-rapid categorization

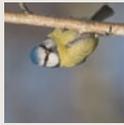
Poncet M.^{1,2,3}, Reddy L.^{1,2}, Fabre-Thorpe M.^{1,2}

¹Université de Toulouse, UPS, Centre de Recherche Cerveau et Cognition, France ; ²CNRS, CerCo, Toulouse, France ; ³Université Toulouse II le Mirail, PDPS, France
 email: poncet@cerco.ups-tlse.fr; lreddy@cerco.ups-tlse.fr; mtf@cerco.ups-tlse.fr



Background

Animal?
 (superordinate-level categorization)



Bird?
 (basic-level categorization)

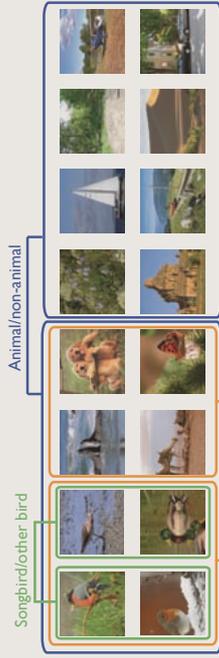
Songbird?
 (subordinate-level categorization)

- Bird accessed first (Rosch et al., 1976; Tanaka & Taylor, 1991).
- Animal accessed first (Mace et al., 2009).
- Animal accessed first only at short stimulus duration (Mack & Palmeri, 2011): fast presentations could degrade the content of the image and emphasize coarse visual information

Which is faster?
 Does it depend on stimulus duration?

Method

3 levels of categorization
 animal/bird/songbird

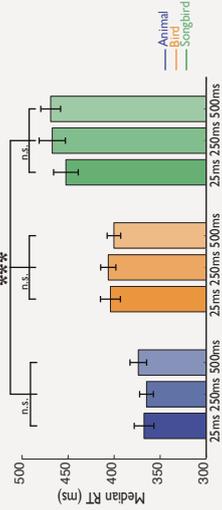


« Release your finger as fast as you can if you see an animal/ bird/ songbird »

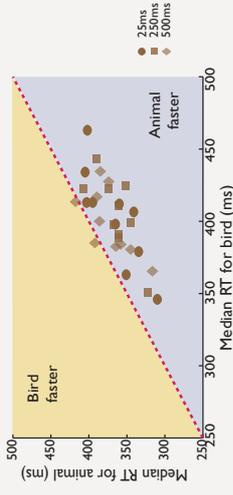
n=10
 200 trials/condition/block (9 blocks)
 3600 images used (no repetition)

Superordinate categorization is faster than basic by 35ms, irrespective of stimulus duration

Median reaction times (RT)

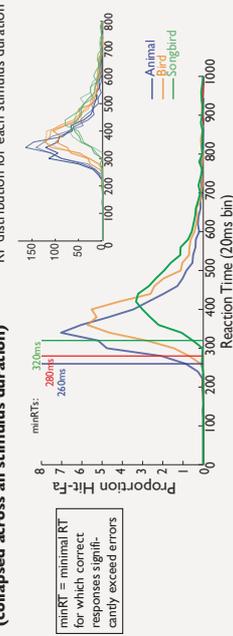


Individual median RT

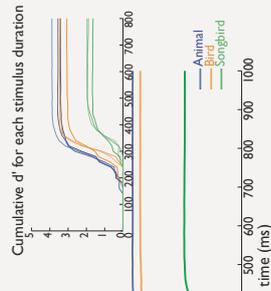


- For all subjects: RT animal < bird << songbird
- Within category, performance does not depend on stimulus duration
- Same results obtained with a 'yes/no task at a stimulus duration of 250ms

RT distribution (collapsed across all stimulus duration)

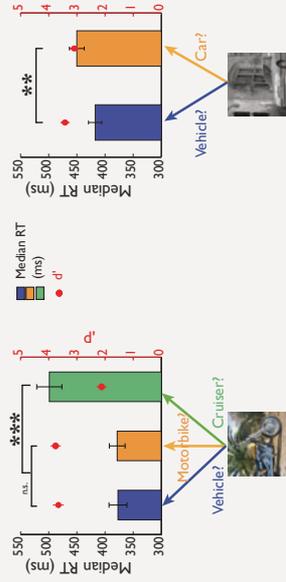


Cumulative d' (collapsed across all stimulus duration)



- Fastest reliable categorization: animal at 260ms, bird at 280ms, songbird at 320ms
- Accuracy for animal > bird >> songbird
- True for the entire response distribution

What about man-made objects?



- No difference between vehicle and motorbike categories
- But car category is accessed 35ms later than vehicle
- Subordinate categorization (cruiser) accessed last

Discussion

- The first category to be accessed is the superordinate-level, regardless of stimulus duration
- At 260ms we are able to reliably categorize an animal but not a bird.
- However, at the same latency, we are able to categorize a vehicle as well as a motorbike
- This might not be true for all man-made categories (car)
- How do we reconcile the current findings with the ones of Mack & Palmeri (VSS 2011)?
 - Diversity of exemplars among targets and distractors?
 - Stimuli presented in block design?
 - Rapidity of responses?

References
 Rosch, E. Marvin, C.B., Gray WD, Johnson DH, Boyes-Braem P. (1976) Cognitive Psychology
 Mace IJ, Jobarrat OK, Napolitano JL, Fabre-Thorpe M. (2009) PLoS ONE
 Palmeri TL, Gauthier GM, Tarr PI, Tanaka JW, Sengco JA, Taylor CS
 Taylor MW & Tarr PI (1991) Cognitive Psychology
 Acknowledgements
 This work was supported by the CNRS and a PhD grant to MP from Toulouse University

2.3. Discussion

Les résultats de cette étude montrent que le temps de présentation n'a aucun effet sur la rapidité de la catégorisation d'objets. Les participants étaient plus rapides et plus précis pour catégoriser un animal au niveau superordonné qu'au niveau basique, et leurs performances étaient les plus faibles pour les catégorisations subordonnées. Ces résultats appuient l'hypothèse d'un traitement des stimuli de manière globale puis de façon plus détaillée (« coarse to fine ») (Navon, 1977; Schyns & Oliva, 1994). Pourtant, même si la décision (ou les temps de réponse) est plus rapide au niveau superordonné, nous ne sommes pas en mesure d'affirmer que les catégories superordonnées sont traitées avant les catégories basiques au niveau neuronal. Cette question est discutée dans le chapitre 2.3.1.

Les résultats obtenus avec la catégorie véhicule sont moins marqués. Nous obtenons un petit avantage des temps de réaction de la catégorie véhicule par rapport à la catégorie voiture mais la catégorisation moto vs. autre véhicule est aussi rapide et précise que la catégorisation véhicule vs. non-véhicule. Notre hypothèse est que la diversité des catégories influencerait les performances de catégorisation (voir la discussion générale de l'article 1). Le seuil de décision pourrait être ajusté en fonction du nombre de caractéristiques permettant de différencier une catégorie d'une autre. Au niveau superordonné peu de caractéristiques sont nécessaires et les temps de réponse sont plus rapides que pour des catégorisations basiques. Cette idée est en accord avec le modèle PDP de McClelland & Rogers, (2003) (décrit dans l'introduction chapitre 3.1) et est plus développée dans le chapitre 2.3.2.

En plus du rôle potentiel du temps de présentation très courts des stimuli sur les performances de catégorisation (qui pourrait être à l'origine d'un avantage des catégorisations superordonnées par rapport aux catégorisations basiques), une autre critique a aussi été émise à propos du paradigme utilisé. Les participants réalisaient des tâches de catégorisation par bloc de 200 essais (un bloc par niveau de catégorisation). La répétition de la même tâche dans un bloc pourrait faciliter l'accès aux catégories d'objets superordonnées alors même que sans répétition, cet accès serait plus lent que pour les catégories basiques. Cette critique est abordée dans le chapitre 2.3.3.

2.3.1. Quel niveau de représentation est accédé en premier ?

Nos résultats montrent que nous sommes plus rapides pour savoir qu'une scène naturelle contient un animal que pour savoir qu'elle contient un oiseau. Par contre, ils ne nous permettent pas de savoir si l'accès aux catégories superordonnées est une étape nécessaire avant l'accès aux autres niveaux de catégories (Mack & Palmeri, 2011a). Les études pouvant réellement nous renseigner sur une différence de temporalité au niveau neuronal, sont celles menées en électrophysiologie. Cependant, leurs résultats sont contradictoires.

Une première étude chez le macaque réalisée par Sugase, Yamane, Ueno, & Kawano, (1999) a montré que les informations globales (différence entre visage humain et visage de macaque) seraient transmises au début de la réponse des neurones de IT. Les informations plus fines (identité ou expression du visage) seraient transmise environ 50 ms plus tard. La même équipe a publié d'autres résultats en faveur d'un traitement d'abord global des stimuli visuels avant un traitement plus fin, mais cette fois en utilisant une analyse par composantes principales des réponses des populations de neurones dans IT (Matsumoto, Okada, Sugase-Miyamoto, Yamane, & Kawano, 2005).

Au contraire, les résultats d'études utilisant des techniques de MVPA (techniques décrites dans le chapitre 1.1.4.1), sont plutôt en faveur d'une apparition simultanée des informations sur la catégorie d'un objet et celles sur son identité. Dans l'étude de Hung et al. (2005), les performances de classification étaient au-dessus du niveau de la chance à partir de 80 ms après l'apparition du stimulus que ce soit pour identifier un stimulus par rapport aux 77 autres ou pour décoder sa catégorie parmi les 8 catégories basiques possibles. Chez l'humain, l'analyse d'enregistrements EEG et MEG a montré que les informations sur la catégorie d'un objet à différents niveau de catégorisation (e.g. animé/non-animé, visage/corps) commencent à apparaître au même moment que les informations sur son identité (Carlson et al., 2013; Cichy, Pantazis, & Oliva, 2014). Ces deux études montrent aussi que les valeurs maximales de classification sont observées pour les informations spécifiques à un stimulus avant les informations catégorielles. Cichy et al. (2014) proposent que différents niveaux de catégorisation atteindraient une classification maximale à différents temps suivant l'accumulation de preuves nécessaires à ces catégorisations.

Les résultats des études venant d'être citées sont difficiles à interpréter et les analyses utilisées sont complexes. Dans les cinq études, les participants (macaques ou humains) n'étaient pas impliqués dans une tâche de catégorisation et on ne sait pas exactement si les résultats obtenus correspondent à des informations qui pourraient être utilisées dans ce type de tâche. Par exemple, que ce soit pour les analyses en composantes principales ou en MVPA, nous ne savons pas quelles sont les caractéristiques que l'algorithme mathématique utilise pour pouvoir séparer deux catégories d'objets ou d'exemplaires. Ces informations pourraient en fait ne pas du tout être utilisées dans les tâches de catégorisation. Dans ce cas, pouvons-nous vraiment parler de reconnaissance d'objet? En effet, des différences de bas niveau (e.g. contraste, luminosité) pourraient permettre à un algorithme de différencier deux objets mais le cerveau humain utilise-t-il ces informations et aussi rapidement? En fait, dans leur article Carlson et al. signalent que : « Given the short onset and peak latencies, within-category exemplar discriminability is likely based on feature representations in early visual areas ». Les résultats de Sugase et al., (1999) même s'ils ne sont pas analysés avec des techniques MVPA se heurtent aux mêmes problèmes. Ils ont utilisé un modèle de transmission d'information mais rien n'indique que ce modèle est correct sur le plan biologique.

Un autre aspect de ces études à prendre en considération est que le nombre de stimuli utilisé est en réalité très faible par rapport au nombre d'exemplaires existant dans notre environnement. Je ne suis pas en train de dire qu'il faille présenter 1 million d'images aux participants humains ou macaques. Ce que j'aimerais signaler ici c'est que l'hyperplan construit par le classifieur pourrait ne pas bien différencier deux catégories d'objets au niveau superordonné puisque très peu d'exemplaires lui sont fournis. Il aurait donc plus de difficultés à repérer les similarités et les distinctions entre, par exemple, les catégories animées et inanimées que pour différents visages (dans ce cas, seulement peu de caractéristiques ou dimensions sont pertinentes).

2.3.2. Importance de la discriminabilité dans la catégorisation

La catégorisation d'objet semble être influencée par la diversité des stimuli au sein des catégories d'objets utilisées. Nos résultats vont encore plus loin puisqu'ils montrent que l'augmentation de la

diversité des images cibles augmente les temps de réaction alors que l'augmentation de la diversité des distracteurs diminuent la précision (d').

L'influence de la diversité des stimuli a aussi été observée dans des tâches de catégorisations de visages (D'Lauro, Tanaka, & Curran, 2008). Des participants étaient testés dans une tâche de vérification et devaient catégoriser des visages au niveau superordonné, basique et subordonné. Les auteurs ont montré qu'en augmentant la similarité perceptuelle des visages (rendant la discrimination plus difficile) la catégorie basique était la plus rapidement exécutée. Au contraire, si les visages distracteurs étaient variés, ils obtenaient un avantage de la catégorie subordonnée. Cet avantage de la catégorie subordonnée pour les visages avait été trouvé et expliqué auparavant par Tanaka, (2001). Selon lui, le point d'entrée de la reconnaissance des visages est déplacé vers la catégorie subordonnée étant donné que c'est une catégorie pour laquelle nous sommes experts (comme c'est le cas pour des experts en oiseaux par exemple, Tanaka & Taylor, 1991). En réalité il semble bien que le choix des distracteurs joue un rôle très important dans les tâches de catégorisation (comme cité auparavant, chien/voiture est beaucoup plus rapide que chien/chat, Bowers & Jones, 2008) et que ce choix a peut-être été pris trop à la légère dans les premières études testant la rapidité de catégorisation des visages (d'autres critiques pourraient être faites à ces deux articles, comme l'utilisation de seulement huit visages à reconnaître présentés au début de l'expérience et qui pourraient donc être stockés en mémoire). On serait plutôt des experts quant au choix des caractéristiques importantes nous permettant de catégoriser les objets en fonction des distracteurs qui leur sont opposés.

Les résultats de notre étude montrent que les performances de reconnaissance dépendraient de la diversité des cibles et des distracteurs. Ces résultats ressemblent à ceux trouvés dans les paradigmes de recherche visuelle (ce paradigme est présenté en détails dans le chapitre 3.1) où il a été montré que la difficulté à trouver un objet visuel cible au sein d'autres objets visuels augmentait avec l'augmentation de la similarité entre la cible et les distracteurs mais aussi avec l'augmentation de la diversité des distracteurs (Duncan & Humphreys, 1989). Il y a déjà plusieurs années, Schyns, (1998) proposait que les performances de reconnaissance étaient le résultat de l'interaction entre les contraintes de la tâche et l'information fourni par le stimulus. Ainsi dans notre cas, le niveau de diversité des stimuli peut être considéré comme le niveau de difficulté de la tâche : plus les cibles et les distracteurs utilisés sont variés et plus la tâche est difficile.

2.3.3. Effet bloc ?

Un autre paramètre (à part le temps de présentation très court) pouvant faire émerger un avantage de la catégorie superordonnée, est le paradigme en « bloc » qui a été utilisé. Dans les expériences de catégorisation telle que celle de l'article 1, les participants sont testés sur le même niveau de catégorisation (e.g. animal/non-animal) pendant 200 essais avant d'être testés sur une autre tâche de catégorisation (e.g. oiseau/animal non-oiseau). Au contraire, dans les tâches de vérification telle que celle de Rosch et al., (1976), les participants doivent catégoriser les images à un niveau d'abstraction différent à chaque essai.

Cet effet du « contexte » a été testé et présenté lors d'une conférence (Vision Science Society 2011) par Mack & Palmeri (2011b) (voir leur poster annexe 8.2). Ils demandaient à des participants de catégoriser des objets visuels au niveau superordonné, basique ou subordonné et manipulaient le temps de présentation des stimuli ainsi que le contexte de la tâche (soit les essais étaient par bloc, c'est-à-dire que la même tâche était répétée pendant plusieurs essais, soit le niveau de catégorisation était aléatoire d'un essai à l'autre). La conclusion de leur étude était qu'il fallait que les images soient présentées rapidement et en bloc pour obtenir un avantage de la catégorisation superordonnée. Il faut signaler cependant que même dans cette condition ils n'observaient pas un avantage de la catégorisation superordonnée, mais seulement des performances équivalentes entre catégorisation basique et superordonnée.

En contradiction avec ces résultats, nous avons montré dans l'article 1 que le temps de présentation n'avait pas d'influence sur les performances de catégorisation. Dans la discussion générale de l'article, nous développons l'idée qu'un paradigme en bloc permettrait de ne pas avoir recours au langage (qui est nécessaire dans les tâches de vérification) et d'optimiser le système visuel dans le traitement des stimuli à un niveau de catégorisation particulier.

Mack & Palmeri (2011b) ont testé plus particulièrement l'effet de la répétition dans l'expérience 4 de leur poster et montré que les temps de réaction diminuaient avec la répétition d'une tâche de catégorisation superordonnée mais que ce n'était pas le cas lorsqu'une tâche de catégorisation basique était répétée (dans ce cas les temps de réponses restent assez stables). Cependant, il est important de noter que ces résultats correspondent à l'analyse des essais répétés au sein de l'expérience. Ainsi, le

premier temps de réaction pour la catégorisation basique est enregistré après une catégorisation superordonnée, et le premier temps de réaction pour la catégorisation superordonnée après une catégorisation basique (sinon on serait déjà dans le cas de la répétition de la même tâche).

L'hypothèse qu'on peut émettre est qu'il est plus facile de passer d'une catégorisation basique à une catégorisation superordonnée que d'une catégorisation superordonnée à une catégorisation basique (voir par exemple les résultats de Evans et al., 2011 montrant un effet de la catégorie cible sur une tâche ultérieure). Le changement de tâche pourrait être coûteux dans un cas (passer du local « chien », au global « animal ») mais pas dans l'autre (passer du global « animal », au local « chien »). Ainsi dans leur expérience, les premiers temps de réaction sont plus courts pour la tâche de catégorisation basique que pour la tâche de catégorisation superordonnée puisque dans le premier cas il n'y a pas d'interférence de l'essai précédent (catégorisation superordonnée sur catégorisation basique) mais qu'il y en a un dans l'autre (catégorisation basique sur catégorisation superordonnée). Cet effet d'interférence diminuerait ensuite lorsque plusieurs catégorisations au niveau superordonné sont répétées.

Pour résumé, on observerait un avantage des catégorisations basiques lorsque les participants doivent souvent changer de tâche puisqu'au contraire des catégorisations superordonnées, elles ne souffrent pas d'un effet d'interférence. Cet effet d'interférence pourrait s'expliquer par une difficulté de passer du local vers le global.

Cette hypothèse peut être testée dans une nouvelle expérience dans laquelle on comparerait les temps de réponse pour les catégorisations basiques après une catégorisation subordonnée. On se trouve alors dans le même cas que pour une catégorisation superordonnée précédée d'une catégorisation basique : un changement de tâche du local vers le global. Si on observe le même pattern de résultats que pour les catégorisations superordonnées, c'est-à-dire des temps de réponse plus lents après une catégorisation subordonnée et qui s'accélèrent avec la répétition de la tâche au niveau basique alors l'effet d'interférence est probablement dû au changement de tâche. Si ce n'est pas le cas, alors une alternative possible serait que, comme dans la vie quotidienne les catégorisations superordonnées sont moins fréquentes que les catégorisations basiques, elles seraient plus vulnérables au changement de tâche.



Chapitre 3
Attention



3. Rôle de l'attention dans la catégorisation d'objets visuels

L'objet de ce chapitre est de déterminer le rôle de l'attention dans l'accès aux différents niveaux de catégories des objets. L'attention a été principalement étudiée dans des paradigmes de recherche visuelle impliquant des stimuli simples, non écologiques (des objets qui n'existent pas dans notre environnement). Récemment des stimuli naturels, tels que des images d'animaux ou de véhicules ont récemment été testés (chapitre 3.1). Dans le précédent chapitre (en particulier dans l'article 1), nous avons montré que toutes les catégorisations d'objets ne sont pas réalisées à la même vitesse (les catégorisations superordonnées étant les plus rapides, suivies par les catégorisations au niveau basique puis au niveau subordonné). L'hypothèse principale qui est testée dans l'article 2 (chapitre 3.2) est que les catégorisations au niveau basique pourraient nécessiter des ressources attentionnelles, au contraire des catégorisations superordonnées. Les résultats de l'article 2 seront ensuite discutés dans le chapitre 3.3.

3.1. L'étude de l'attention

Le rôle de l'attention dans la reconnaissance d'objets a été principalement étudié dans des expériences utilisant un paradigme de recherche visuelle (revue M. P. Eckstein, 2011). Différentes variantes de ce paradigme existent mais dans une tâche de recherche visuelle standard, les participants cherchent un stimulus cible parmi des distracteurs (Figure 18, panel de gauche). Le dispositif expérimental est présenté jusqu'à la réponse du participant qui doit indiquer si la cible est présente ou absente (en général elle est présente dans 50% des essais). Le nombre de distracteurs varie suivant les essais. Etant donné que les performances en termes de réponses correctes sont très élevées, on s'intéresse au temps de réaction pour trouver la cible en fonction du nombre de distracteurs (Figure 18, panel de droite). La pente d'accroissement (ou coefficient directeur) est appelée pente de recherche et est supposée refléter la demande en attention. Si tous les stimuli peuvent être traités en une seule fois à un niveau suffisant pour distinguer la cible des distracteurs, alors la pente de recherche est proche de zéro (recherche parallèle). Au contraire, si les stimuli demandent chacun de l'attention pour être traité, alors les temps de réaction augmentent avec l'augmentation du nombre de distracteurs (recherche sérielle). La pente de recherche obtenue correspond à une recherche aléatoire (par chance le premier stimulus regardé est la cible mais parfois il est le dernier), c'est-à-dire qu'en moyenne l'attention doit être focalisée sur la moitié des stimuli avant la réponse du participant.

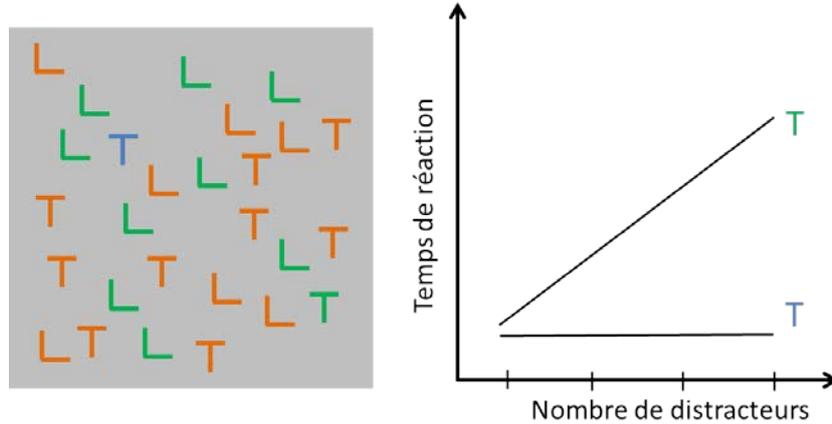


Figure 18. Exemple d'une tâche de recherche visuelle (panel de gauche). Les participants doivent chercher soit un T bleu, soit un T vert. Dans le cas du T bleu, la cible se distingue par un trait visuel particulier, la couleur bleue. Quel que soit le nombre de distracteurs, le temps de réponse ne varie pas (panel de droite). La pente de recherche reste stable, ce qui définit une recherche parallèle. Dans le cas de la recherche du T vert, conjonction des traits 'forme T' et 'couleur verte', les temps de réponse augmentent avec le nombre de distracteurs. Ce type de recherche visuelle est dite sérielle.

3.1.1. Des stimuli simples...

Les travaux pionniers de Treisman et Julesz ont largement influencé le domaine de la vision (e.g. Bergen & Julesz, 1983; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman & Souther, 1985). Leurs résultats montrent que les tâches de recherche visuelle où la cible peut être distinguée des distracteurs par un trait visuel unique comme la couleur, la taille, le mouvement, ou l'orientation, sont des recherches réalisées par des mécanismes capables de traiter les stimuli en parallèle. Tous les autres types de recherches visuelles, incluant les recherches de conjonction de traits (par exemple chercher un T vert parmi des T oranges et des L verts) seraient basés sur des mécanismes qui opèrent en série, stimulus par stimulus.

Pour expliquer la différence entre ces deux types de recherche visuelle Treisman & Gelade, (1980) ont proposé une théorie d'intégration des traits visuels (« Feature-integration theory »). D'après cette théorie, la perception visuelle du monde serait basée sur l'extraction en parallèle d'un set limité de traits élémentaires pré-attentionnels. L'attention serait alors nécessaire pour rassembler ces traits visuels en un unique objet (processus de « binding »).

Un des arguments montrant que l'attention est nécessaire pour unir plusieurs traits visuels d'un même objet est l'existence de conjonctions illusoires (Treisman & Schmidt, 1982). Ce type d'illusion consiste en la perception d'un nouvel objet résultant de l'association de traits élémentaires de différents objets. Par exemple, si un X vert et un O rouge sont flashés sur un écran si brièvement que l'attention ne puisse pas intervenir, les participants disent parfois avoir vu un X rouge et un O vert. Dans ce cas, les traits élémentaires ont été correctement extraits de la scène puisqu'ils ne requièrent pas d'attention par contre, leur intégration à un objet en particulier n'a pas pu s'effectuer correctement car l'attention n'a pas eu le temps de jouer ce rôle.

Bien que la théorie de l'intégration des traits ait beaucoup influencé et stimulé la recherche dans ce domaine, une stricte dichotomie entre les tâches de recherche parallèle et sérielle ne peut pas expliquer tous les résultats obtenus par la suite (voir par exemple les revues de Chun & Wolfe, 2001; Nakayama & Joseph, 1998; Wolfe, 1998 ; ainsi que la Figure 19). Il semblerait plutôt que les performances (pentes de recherche) des participants se situent sur un continuum entre ces deux extrémités. On parlera alors de recherche plus ou moins efficace. La variabilité de l'efficacité des recherches visuelles a été expliquée (entre autres) par les modèles de recherche guidée. Elle serait déterminée par l'habileté des traitements pré-attentifs ou parallèles à guider l'attention vers une cible potentielle ou à l'éloigner des distracteurs (Cave & Wolfe, 1990; Wolfe, Cave, & Franzel, 1989; Wolfe, 1994, 2003).

<i>Factors that decrease search efficiency</i>	<i>Factors that increase search efficiency</i>
In general, as target-distractor differences get smaller, search becomes less efficient (e.g., Foster & Westland, 1992; Nagy & Sanchez, 1990).	Large target-distractor differences in features such as color, orientation, motion, size, curvature, some other form properties, and some 3-D properties (such as stereopsis, lighting, and linear perspective). See Wolfe (1998b) for a review.
Increasing distractor inhomogeneity. Consult Duncan and Humphreys (1989) for a detailed discussion of the role of similarity in visual search.	Increasing distractor homogeneity (Duncan, 1988).
Targets defined by conjunctions of two or more basic features (Treisman & Gelade, 1980; e.g., color X orientation: a red vertical line among green vertical and red horizontal distractors).	Conjunction targets can be found efficiently if the differences in target and distractor features are sufficiently salient (Wolfe, Cave, & Franzel, 1989).
Targets defined only by the spatial arrangement of basic features are, as a general rule, not found efficiently (Wolfe & Bennett, 1997). Thus, search for an "S" among mirror-reversed Ss will proceed at a rate of 20–30 msec per item on target present trials.	Difficult searches can become more efficient with extensive practice (Heathcote & Mewhort, 1993; Treisman, Vieira, & Hayes, 1992). However, such perceptual learning is specific to the training stimuli.

Figure 19. Facteurs influençant l'efficacité des tâches de recherche visuelle. (Wolfe, 1998)

En résumé, la détection de traits basiques ou pré-attentifs pourrait être réalisée en parallèle alors que la détection de stimuli plus complexes demanderait l'examen de chaque stimulus par une sorte de faisceau attentionnel. L'attention serait donc une condition nécessaire à la reconnaissance d'objets puisqu'elle est décrite comme le mécanisme permettant de relier différents traits visuels élémentaires en un objet (Treisman & Kanwisher, 1998; Treisman, 1996). Cela implique qu'aucune reconnaissance d'objet complexe ne peut avoir lieu en dehors du focus attentionnel. Qu'en est-il vraiment ?

3.1.2. ... aux stimuli complexes

Les premiers travaux utilisant des formes complexes dans des tâches de recherche visuelle ont montré que même si plusieurs traits visuels d'un stimulus peuvent être représentés pré-attentivement, l'attention est nécessaire pour connaître les relations entre ces traits (Wolfe & Bennett, 1997). Les auteurs ont testé dans plusieurs expériences divers stimuli complexes (conjonction de traits et couleur, formes arbitraires, formes ressemblant à des poules, etc..) et ont observé que même si les participants semblaient savoir pré-attentivement qu'un objet visuel possédait certains attributs (rouge, ligne

verticale), l'attention était nécessaire pour déterminer les relations entre ces attributs (quelle partie de l'objet était rouge *et* verticale). Quelques informations sur la structure d'un objet seraient donc disponibles pré-attentivement mais la représentation de la forme globale d'un objet nécessiterait l'intégration de traits visuels locaux, un mécanisme qui demande de l'attention.

Pourtant, plusieurs études ont commencé à suggérer que les choses pourraient changer si des stimuli naturels étaient utilisés à la place de stimuli abstraits. Par exemple, Rousselet et al. (2002) ont montré que la catégorisation de scènes naturelles pouvait être réalisée en parallèle (sans que l'attention focale ne soit nécessaire). Les participants de cette expérience voyaient simultanément soit une ou deux scènes naturelles et devaient répondre si l'image (ou une des deux images) contenait un animal. Les performances entre les deux conditions (présentation d'une ou de deux images) étaient très similaires en termes de précision, de temps de réponse ou même en considérant les potentiels évoqués (enregistrés en EEG).

Utilisant un autre type de paradigme, mais toujours une tâche de catégorisation animal/non-animal avec des scènes naturelles, les résultats de F. F. Li, VanRullen, Koch, & Perona, (2002) suggèrent aussi que ce type de catégorisation pourrait être réalisé en quasi-absence d'attention. Les auteurs ont utilisé un paradigme de double tâche (paradigme détaillé dans l'article 2). Les participants devaient effectuer une tâche centrale sollicitant toutes ou la majorité de leurs ressources attentionnelles (Braun & Julesz, 1998; Braun, 1994) pendant que des scènes naturelles étaient flashées en périphérie. Les performances des participants dans les catégorisations animal/non-animal et véhicule/non-véhicule étaient similaires que la tâche soit réalisée seule (attention disponible) ou simultanément avec la tâche centrale qui devait être réalisée en priorité (la majorité des ressources attentionnelles est focalisée sur la tâche centrale). Au contraire, s'ils devaient réaliser une tâche périphérique connue pour être attentionnelle (discrimination d'un disque vert-rouge d'un disque rouge-vert, Braun & Julesz, 1998; Braun & Sagi, 1990; Lee, Koch, & Braun, 1999), leurs performances étaient très détériorées en condition de double tâche, par rapport à la condition où la tâche périphérique était réalisée seule.

Ce paradigme de double tâche a été réutilisé pour valider ces résultats et tester la demande en attention pour d'autres stimuli. Il a ainsi été montré qu'il était possible de détecter la présence d'un animal dans une scène naturelle en quasi-absence d'attention (les performances n'étaient pas diminuées en condition de double tâche par rapport à la tâche réalisée seule) même si les scènes étaient en niveaux

de gris ou si deux images étaient présentées en périphérie simultanément (Fei-Fei, VanRullen, Koch, & Perona, 2005). Nous sommes aussi capables de reconnaître le genre des visages (mâle vs. femelle, Leila Reddy, Wilken, & Koch, 2004) mais aussi l'identité d'un visage célèbre (Leila Reddy, Reddy, & Koch, 2006) en quasi-absence d'attention.

Ces résultats sont surprenants puisque d'après les études utilisant un paradigme de recherche visuelle, aucune reconnaissance d'objet ne pourrait avoir lieu en dehors du focus attentionnel. En fait, si on demande à des participants de chercher un animal parmi d'autres scènes naturelles, les performances diminuent avec l'augmentation du nombre de distracteurs (VanRullen, Reddy, & Koch, 2004). Ces résultats indiquent que la recherche est inefficace et qu'elle demande donc des ressources attentionnelles. Comment réconcilier ces résultats ?

3.1.3. La théorie de l'intégration des traits revisitée

Chercher une scène contenant un animal parmi d'autres scènes naturelles nécessite de l'attention alors que la catégorisation animal/non-animal ne nécessite que peu ou pas d'attention. Pour résoudre ces résultats contradictoires, VanRullen et al. (2004) proposent que les mécanismes attentionnels impliqués dans le paradigme de recherche visuelle et celui de double tâche ne soient pas les mêmes. Les performances obtenues dans un paradigme de recherche visuelle permettraient de distinguer des processus de traitement parallèles vs. sériels alors que celles obtenues dans un paradigme de double tâche distingueraient des traitements attentifs vs. pré-attentifs.

Ces deux mécanismes seraient indépendants et une tâche pourrait alors être sérielle et pré-attentive comme c'est le cas pour les tâches de catégorisation d'objets, mais toutes les combinaisons sont possibles (Figure 20A). Par exemple il existe également des tâches pré-attentives et parallèles (e.g. discriminations d'orientation ou de couleur ; Braun & Julesz, 1998; Treisman & Gelade, 1980), des tâches attentives et sérielles (e.g. L vs.T dans des rotations aléatoires, disque vert-rouge vs. rouge-vert ; Bergen & Julesz, 1983; Lee et al., 1999; VanRullen et al., 2004) ou des tâches attentives et parallèles (e.g. L vs.+ dans des rotations aléatoires, discrimination de cubes; Bergen & Julesz, 1983; Enns & Rensink, 1990; Sun & Perona, 1996; VanRullen et al., 2004).

Pour expliquer la différence entre processus parallèles et pré-attentifs, les auteurs émettent les hypothèses que (Figure 20):

- un trait, un stimulus ou une catégorie d'objet est *pré-attentif* s'il existe une population de neurones sélectifs à ce trait, stimulus ou catégorie d'objet, indépendamment des aires corticales impliquées. Cette hypothèse est basée sur l'existence de neurones sélectifs à l'orientation et à la couleur, deux dimensions qui sont clairement pré-attentive (e.g. Braun & Julesz, 1998) mais il existe aussi des neurones sélectifs à des catégories particulières d'objets comme des animaux ou des visages (e.g. Gross et al., 1972; Kreiman, Koch, & Fried, 2000). De plus, la sélectivité de ces neurones est préservée chez les animaux anesthésiés (Perrett et al., 1982) donc l'attention et d'autres facteurs top-down ne joueraient qu'un rôle mineur dans l'activité de ces neurones sélectifs (Desimone et al., 1984).
- les stimuli détectés en *parallèle* bénéficieraient de traitements ayant lieu dans les aires visuelles primaires (V1-V2). Les champs récepteurs étant petits dans ces régions, cela permettrait d'isoler facilement les différents stimuli et donc de détecter les stimuli en parallèle.

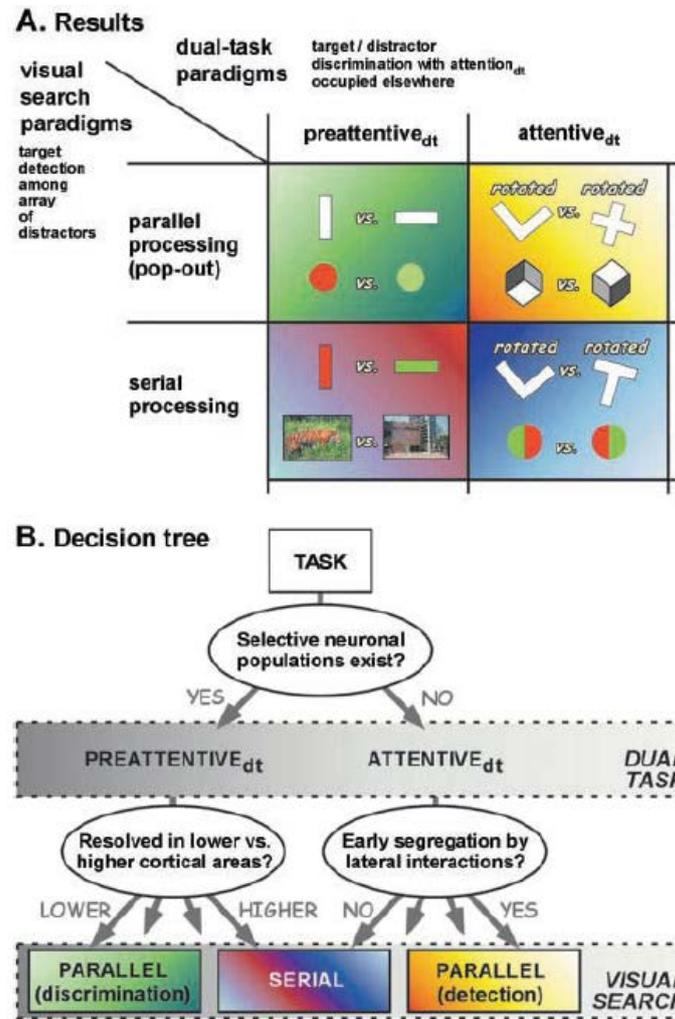


Figure 20. Hypothèses pour expliquer la différence entre l'interprétation des résultats d'un paradigme de recherche visuelle ou d'un paradigme de double tâche. A. Résultats obtenus pour différents stimuli testés dans des tâches de recherche visuelle et de double tâche. La recherche visuelle teste si le traitement d'un type de stimulus se fait en parallèle ou de façon sérielle alors que la double tâche teste si une tâche peut être réalisée sans attention. B. Une tâche est pré-attentive s'il existe des populations de neurones qui répondent sélectivement aux stimuli cibles et distracteurs, sinon elle demande de l'attention pour être réalisée. Le traitement en parallèle de plusieurs stimuli dépendrait, lui, du lieu de traitement des stimuli. Si il est réalisé dans les aires visuelles primaires alors la pente de recherche est nulle, si il est réalisé dans des aires plus tardives alors la pente de recherche est sérielle. (VanRullen et al., 2004)

3.1.3.1 Un mécanisme d'intégration des traits automatique

Même si on accepte les hypothèses proposées par VanRullen et al. (2004) de deux mécanismes attentionnels différents pour les paradigmes de recherche visuelle et de double tâche, ces hypothèses ne

résolvent pas le problème de l'intégration des traits visuels qui est supposé être le produit d'un mécanisme attentionnel (Treisman & Gelade, 1980). En fait, les résultats obtenus pour les objets complexes suggèrent l'existence d'une intégration des traits visuels élémentaires ne nécessitant pas d'attention.

On pourrait alors supposer qu'un mécanisme d'intégration soit codé au sein du traitement des informations visuelles (Rousselet, Thorpe, & Fabre-Thorpe, 2004). Avec l'expérience (tout au long de la vie), des neurones de V4 pourraient avoir appris à répondre à certaines configurations de pattern d'activité. Ces neurones seraient encore sélectifs à la localisation spatiale (au moins partiellement) mais leur activité permettrait de déterminer rapidement la présence d'un objet. Une dernière étape optionnelle pourrait ensuite fusionner les réponses des neurones de V4 partageant des sélectivités similaires pour produire une réponse invariante spécifique à un objet. Un tel système serait alors capable de traiter plusieurs objets familiers simultanément.

L'hypothèse d'un mécanisme d'intégration automatique des traits élémentaires pour les catégories d'objets naturels et familiers a été testée dans deux expériences (Lavanya Reddy & VanRullen, 2007; VanRullen, Reddy, & Fei-Fei, 2005). Ces expériences reposent sur l'idée qu'un tel mécanisme d'intégration automatique existe seulement pour a) des catégories de stimuli familiers souvent rencontrés dans notre environnement et b) des objets qui sont assez éloignés spatialement les uns des autres (sinon les configurations apprises se superposeraient dans le même champ récepteur).

Dans une première expérience (VanRullen et al., 2005), les auteurs ont utilisé un paradigme de double tâche similaire à F. F. Li et al., (2002). Les participants devaient réaliser une tâche de discrimination centrale de lettres et/ou une tâche périphérique de catégorisation animal/non-animal. En plus de la scène naturelle présentée en périphérie (cible), une autre scène naturelle que les participants devaient ignorer (distracteur) était présentée simultanément, soit près de la cible, soit éloignée. La présence du distracteur diminuait les performances en double tâche mais seulement dans le cas où il était proche de la cible. L'interférence entre cible et distracteur se produisait donc seulement au sein d'un voisinage local de la cible. Les auteurs proposent que lorsque les deux objets sont présentés trop près l'un de l'autre (comme c'est le cas dans une recherche visuelle), les mécanismes automatiques d'intégration des traits visuels ne pourraient pas jouer leur rôle et la tâche devrait alors être réalisée par

une intégration des traits visuels nécessitant de l'attention (comme proposé par Treisman & Gelade, 1980).

Dans une seconde expérience (Lavanya Reddy & VanRullen, 2007), les auteurs ont testé des participants dans deux tâches de recherche visuelle: visages à l'endroit vs. à l'envers et disque rouge-vert vs. vert-rouge. Les auteurs faisaient varier le nombre de distracteurs et la distance entre les stimuli. Avec l'augmentation de la distance entre les stimuli, les performances des participants s'amélioraient pour la tâche de discrimination de visages tandis qu'aucun effet n'était observé pour la discrimination de disques. La recherche visuelle serait sérielle pour des raisons différentes dans les deux types de stimuli utilisés. Dans le cas de la discrimination de disque (stimuli attentionnels), la recherche serait sérielle parce que la tâche nécessiterait l'intégration de traits visuels par l'attention. Dans le cas de la discrimination de visages (stimuli pré-attentifs), la recherche serait sérielle car il y aurait une compétition au niveau neuronal entre la cible et les distracteurs. L'intégration des traits visuels sans attention ne pourrait pas être réalisée correctement du fait du chevauchement de traits visuels appartenant à deux images différentes dans le champ récepteur des neurones.

Ainsi, deux mécanismes d'intégration des traits visuels coexisteraient : un mécanisme permettant de détecter certaines configurations familières de manière automatique et un mécanisme nécessitant de l'attention pour des stimuli non familiers ou si des configurations familières sont présentées trop près les unes des autres (VanRullen, 2009).

3.1.3.2 Nouveau modèle de recherche visuelle

Basé sur le modèle de recherche guidée (Wolfe, 2003), un nouveau modèle de la recherche visuelle a été développé (Wolfe, Võ, Evans, & Greene, 2011). Ce modèle suppose l'existence d'une voie non-sélective en plus de la voie de traitement classique (voie sélective où les traits visuels élémentaires seraient intégrés à l'aide de l'attention) (Figure 21). La voie non-sélective permettrait d'accéder à une idée générale d'une scène visuelle, « gist » (Oliva, 2005), sans avoir à identifier chaque partie d'un objet et d'être ainsi capable de catégoriser des scènes flashées très rapidement sur la base d'informations statistiques très globales (contexte naturel vs. urbain : Greene & Oliva, 2009a, 2009b; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007 ; présence d'un animal ou non : Evans & Treisman, 2005;

Kirchner & Thorpe, 2006). Les auteurs du modèle ne font pas référence aux travaux cités précédemment mais cette voie correspondrait approximativement à l'hypothèse de l'existence de mécanismes d'intégration des traits visuels de façon automatique (Rousselet et al., 2004; VanRullen, 2009).

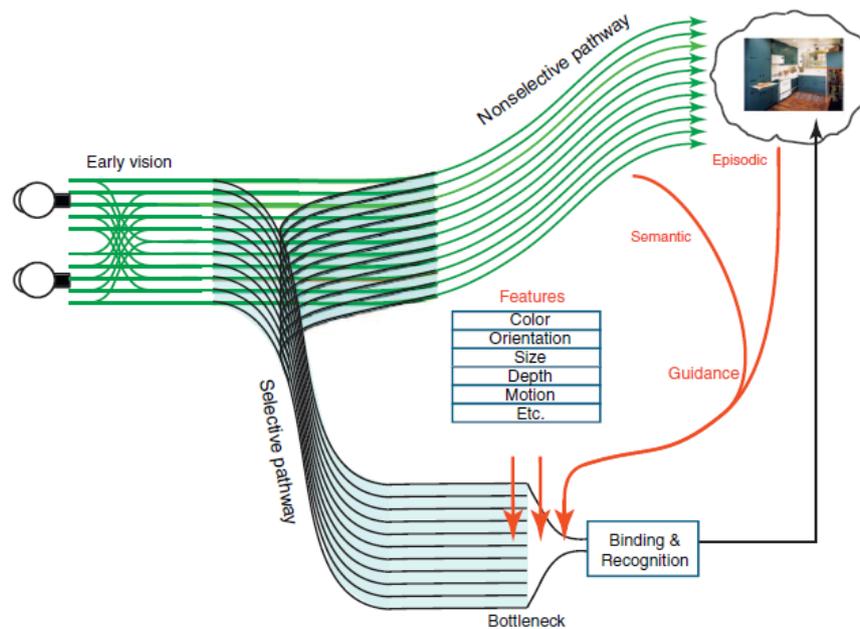


Figure 21. Architecture en deux voies du traitement visuel. La voie non-sélective peut extraire des statistiques de l'image et permettre de sélectionner des cibles potentielles dans l'image par des mécanismes de guidance classiques (e.g. couleur) ou provenant de la scène (sémantique et épisodique). Par contre, elle ne permet pas l'identification précise des objets. Cette étape demande l'intégration des traits visuels d'un objet entre eux par la voie sélective qui a une capacité limitée (« bottleneck »). (Wolfe, Vö, et al., 2011)

Ce modèle prend aussi en compte les résultats obtenus dans diverses expériences, en particulier ceux obtenus dans des tâches de recherches visuelles utilisant des images naturelles (recherche d'un objet dans le contexte d'une scène). Les résultats de ces études montrent que la recherche d'un objet est guidée par le contexte sémantique ou épisodique de la scène. Par exemple, chercher une baguette de pain dans une cuisine est plus rapide que de chercher ce même objet au sein d'autres objet de même type (Vickery, King, & Jiang, 2005; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). Ce

guidage sémantique pourrait être généré par la voie non-sélective et ne serait donc pas soumis aux ressources limitées de l'attention. Une illustration du modèle est présentée Figure 22.

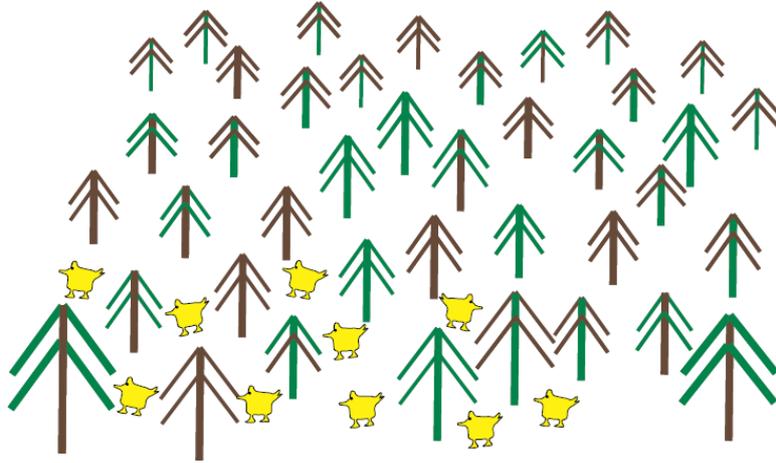


Figure 22. Exemple illustrant le modèle de deux voies de traitement dans la recherche visuelle. La voie non-sélective permettrait d'identifier les statistiques de l'image. Le participant verrait grâce à elle le « gist » d'une forêt et peut-être la présence d'animaux. Cependant l'identification d'un arbre avec un tronc marron et des branches vertes, ou d'un oiseau orienté vers la droite, ferait appel à la voie sélective. (Wolfe, Vö, et al., 2011)

3.1.4. Présentation de l'article 2

Les résultats présentés jusqu'ici suggèrent que les objets complexes pourraient être catégorisés en quasi-absence d'attention. Ces mécanismes reposeraient sur une intégration automatique de patterns de traits visuels familiers (Rousselet et al., 2004) potentiellement grâce à l'existence de neurones sélectifs à certaines catégories de stimuli (VanRullen, 2009). Cependant le niveau de détail des représentations créées par ces mécanismes reste inconnu. Il est suffisant pour reconnaître des visages ou des animaux, mais est-il suffisant pour catégoriser des objets au niveau basique (faire la différence entre un chien et un autre animal) ? En effet, cette voie d'intégration automatique pourrait permettre de traiter seulement des traits diagnostiques d'une scène naturelle et donc de n'accéder qu'au « gist » de la scène (e.g. Wolfe, Vö, et al., 2011). Par ailleurs, nous avons montré dans l'article 1 que les catégorisations basiques requièrent un traitement additionnel par rapport aux catégorisations

superordonnées. Ce temps additionnel pourrait-il s'expliquer par un besoin en attention pour fusionner les traits visuels des objets ?

Pour répondre à ces questions, nous avons testé le niveau de précision des informations que l'on a sur des stimuli situés en dehors du focus attentionnel. Plus spécifiquement, nous avons demandé à des participants de réaliser des catégorisations d'images naturelles au niveau superordonné (animal/non-animal) et basique (chien/animal non-chien) dans un paradigme de double tâche similaire à celui de F. F. Li et al. (2002). Dans une seconde expérience, nous avons testé ces deux niveaux de catégorisation mais en utilisant des objets artificiels (véhicule/non véhicule pour la catégorisation superordonnée, voiture/véhicule non-voiture pour la catégorisation basique). Dans les deux expériences, pour atteindre un même niveau de performance, les stimuli (masqués) devaient être présentés plus longtemps pour les catégorisations au niveau basique que pour les catégorisations superordonnées. Nous avons alors testé dans une troisième expérience le niveau de performance des participants pour ces deux niveaux de catégorisation lorsque les images masquées étaient présentées pendant la même durée avant l'apparition du masque.

3.2. Article 2: A need for more information uptake but not focused attention to access basic-level representations

A need for more information uptake but not focused attention to access basic-level representations

Centre de Recherche Cerveau and Cognition, UPS,
Université de Toulouse, Toulouse, France,
CerCo, CNRS, Toulouse, France, &
PDPS, Université Toulouse II le Mirail,
Toulouse, France

Marlene Poncet



Centre de Recherche Cerveau and Cognition, UPS,
Université de Toulouse, Toulouse, France, &
CerCo, CNRS, Toulouse, France

Leila Reddy



Centre de Recherche Cerveau and Cognition, UPS,
Université de Toulouse, Toulouse, France, &
CerCo, CNRS, Toulouse, France

Michele Fabre-Thorpe



Complex visual scenes can be categorized at the superordinate level (e.g., animal/non-animal or vehicle/non-vehicle) without focused attention. However, rapid visual categorization at the basic level (e.g., dog/non-dog or car/non-car) requires additional processing time. Such finer categorization might, thus, require attentional resources. This hypothesis was tested in the current study with a dual-task paradigm in which subjects performed a basic-level categorization task in peripheral vision either alone (single-task condition) or concurrently with an attentionally demanding letter discrimination task (dual-task condition). Our results indicate that basic-level categorization of either biological (dog/non-dog animal) or man-made (car/non-car vehicle) stimuli requires more information uptake but can, nevertheless, be performed when attention is not fully available, presumably because it is supported by hardwired, specialized neuronal networks.

Keywords: attention, superordinate-level categorization, basic-level categorization, natural scenes, dual task, pre-attentive recognition

Citation: Poncet, M., Reddy, L., & Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision*, 12(1):15, 1–16, <http://www.journalofvision.org/content/12/1/15>, doi:10.1167/12.1.15.

Introduction

We are constantly bombarded by visual information from our environment. Attending to pertinent stimuli enables us to reduce the amount of information to be processed and, consequently, makes visual processing more efficient. However, visual attention has a limited capacity and only a restricted number of stimuli can benefit from attention at a time. What information do we have about objects located outside the attentional focus when attentional resources are directed toward another specific object? Previous studies that explored this question showed that humans are able to perform simple tasks, such as orientation and color discrimination, in the near absence of attention (Braun, 1994; Braun & Julesz, 1998; Braun & Sagi, 1990; Julesz & Schumer, 1981; Treisman & Gelade, 1980). On the contrary, the discrimination of motion or of slightly more complex stimuli composed of conjunctions of simple features, such as randomly oriented Ls and Ts, or spatial arrangements of two colors,

cannot be performed when attention is engaged elsewhere (Lee, Koch, & Braun, 1999).

In contrast to these more artificial types of stimuli, the processing of natural stimuli (e.g., scenes and faces) is remarkably efficient even in the absence of attention. Response latencies at the behavioral and electrophysiological levels show that humans are equally fast at detecting animals when two scenes are presented simultaneously as they are for just one (Rousselet, Fabre-Thorpe, & Thorpe, 2002). Moreover, reporting if a scene contains an animal or a vehicle (Li, VanRullen, Koch, & Perona, 2002) or whether a face is feminine or masculine (Reddy, Wilken, & Koch, 2004) can be done with little drop in behavioral accuracy even when focused attention is engaged elsewhere. To account for these results, it has been proposed that the categorization of frequently encountered stimuli, such as natural scenes and faces, could rely on selective populations of neurons that have developed over time and that could be activated even in the absence of attention. In contrast, the processing of rarely encountered artificial stimuli would be dependent

on the availability of attentional resources (Fei-Fei, VanRullen, Koch, & Perona, 2005; VanRullen, 2009; VanRullen, Reddy, & Koch, 2004). The automatic visual processing of familiar stimuli could, thus, be an alternative mechanism of the brain to effortlessly navigate the ocean of visual information.

The studies mentioned above have demonstrated that natural scene categorization at the superordinate level (e.g., animal/non-animal or vehicle/non-vehicle) is feasible in the near absence of focal attention. However, it is generally recognized that object categorization—grouping different objects into one category and discriminating them from items from another category—can occur at different levels of specificity or detail (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Rosch et al. proposed a distinction between three category levels: superordinate level (e.g., discriminating an animal from a vehicle), basic level (e.g., discriminating a dog from another animal), and subordinate level (e.g., discriminating between a Dalmatian and another dog). Although the prevalent consensus supports the idea of a basic-level advantage (Murphy & Brownell, 1985), recent studies have suggested that access to basic visual categories takes longer processing time compared to superordinate visual categories and that this might occur at different stages of processing. First, more incoming information is required for basic than for superordinate level categorization. Indeed, global properties of an image are available with less presentation time than basic-level details (Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009). Second, access to the basic-level category requires additional processing time. Reaction times are longer for basic than for superordinate-level categorization both for objects (Mace, Joubert, Nespoulous, & Fabre-Thorpe, 2009) and scenes (Joubert, Rousset, Fize, & Fabre-Thorpe, 2007). These behavioral measures are consistent with electrophysiological studies reporting that global information is transmitted earlier than finer grained information (Sugase, Yamane, Ueno, & Kawano, 1999). The delay in accessing more detailed information at the basic level could reflect additional processing of stimulus information, perhaps as a result of feedback signals and/or attentional processing (Ahissar, Nahum, Nelken, & Hochstein, 2009; Cromer, Roy, & Miller, 2010; Freedman & Miller, 2008; Freedman, Riesenhuber, Poggio, & Miller, 2003; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008).

The purpose of the present study was to understand the nature of the additional processing required for basic-level categorization and, more specifically, to determine if the longer processing time could be explained—at least partly—by a necessary allocation of attentional resources. To determine the attentional requirements of basic-level categorization, we used a dual-task paradigm with two types of stimuli often encountered in our everyday life. In the first experiment, participants were tested with a natural

category (i.e., a dog/non-dog categorization task), while in the second experiment they were tested with a man-made category (i.e., a car/non-car categorization task). Participants performed each basic-level categorization task either alone, when attention was fully available (single-task condition), or concurrently with an attentionally demanding task that captured most of (if not all) the available attentional resources (dual-task condition). Categorization performance was then compared in the single- and dual-task conditions. If the basic-level categorization task does not require attention, then performance on this task should be very similar in the single- and dual-task conditions. On the other hand, performance in the dual-task condition should be considerably impaired if attention is necessary for successful categorization at the basic level (Braun & Julesz, 1998; Braun & Sagi, 1990; Fei-Fei et al., 2005; Lee et al., 1999; Li et al., 2002; Reddy et al., 2004). Finally, a third experiment tested the amount of information participants were able to glean from the stimuli depending on its presentation time.

Methods

Participants

Six participants were tested in Experiment I. Four of these and two additional participants were tested in Experiment II. In Experiment III, five participants tested in Experiment I were again tested with biological stimuli, and five tested in Experiment II were tested with man-made stimuli. One author (MP) participated in all three experiments; another author (LR) took part in Experiments I and III. All participants (5 women, 3 men, 1 left-handed, 7 right-handed, between 22 and 35 years of age) had normal or corrected-to-normal acuity and provided written informed consent.

Image database

The images used in the current experiments were chosen from the image sets used by previous studies (Joubert, Fize, Rousset, & Fabre-Thorpe, 2008; Mace et al., 2009; Reddy et al., 2004), from the LabelMe database (Russell, Torralba, Murphy, & Freeman, 2008), and from the Internet. Each image, subtending $5^\circ \times 5^\circ$ of visual angle, was presented in the periphery. They were all grayscale converted to avoid potential facilitation effects of color on recognition (Gegenfurtner & Rieger, 2000; Rossion & Pourtois, 2004; Yip & Sinha, 2002). The images used for the training and testing phases were not the same and each image was viewed at most 6 times by a subject during the entire experiment (3 times in the

superordinate-level categorization task and 3 times in the basic-level categorization task). Participants categorized natural scenes that contained one or several visual objects belonging to one of four categories: animals, vehicles, dogs, and cars (for examples, see [Figure 1](#)). The animal category (non-dog animals, $n = 1384$) contained mammals, birds, fish, reptiles, and insects; the vehicle category (non-car vehicles, $n = 1178$) included trucks, boats, airplanes, motorbikes, bicycles, balloons, etc.; the dog category ($n = 1078$) included dogs of various breeds: Dalmatians, Poodles, Spaniels, Bulldogs, etc.; and, the car category ($n = 830$) included all kinds of cars: racing cars, sport cars, vintage cars, modern cars, and moving, immobile, and damaged cars. The objects present in the images were of various orientations, positions, and sizes. The context surrounding them could be natural or urban, indoor or outdoor.

Apparatus

Participants were seated in a dimly lit room, approximately 75 cm from a computer screen (1024 × 768 pixels, refresh rate: 100 Hz). The stimulus display was synchronized with the refresh rate of the monitor.

Procedure

We used a dual-task paradigm. Participants performed two tasks: a central attentionally demanding task and a peripheral task for which attentional requirements were tested. These two tasks were performed either alone, when attention was fully available (single-task condition), or simultaneously (dual-task condition) with subjects instructed to prioritize the central task and perform the peripheral task as well as they could. If the peripheral task does not require attention, then performance should be very similar in the single- and dual-task conditions. On the other hand, if it requires attentional resources, then performance in the dual-task condition should be considerably impaired.

Dual-task paradigm

The experimental timeline for one trial is illustrated in [Figure 2](#). The stimulus display was the same for all conditions and only the instructions differed for each condition. Each trial started with a fixation cross, followed by the presentation of the central attentionally demanding task (letter discrimination task, see below for more

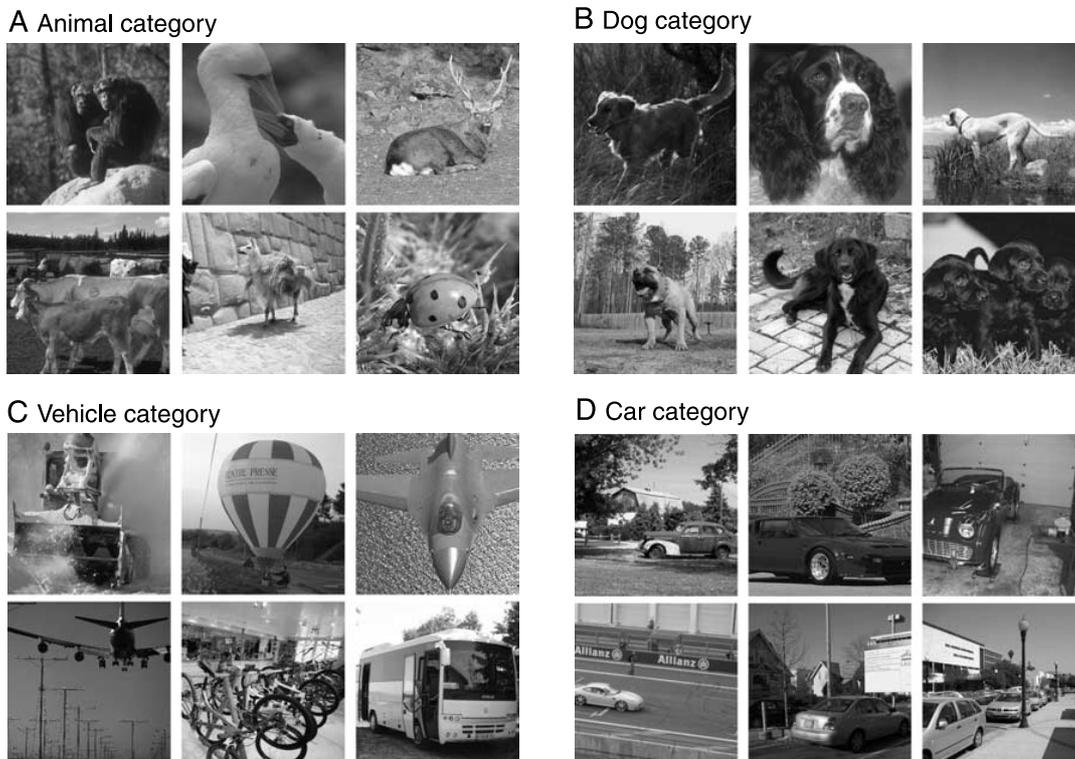


Figure 1. Examples of natural scene images used in the study. In Experiment I (top row), the objects contained in the images were biologically pertinent and could be categorized as (A) animal at the superordinate level or (B) dog at the basic level. In Experiment II (bottom row), stimuli were man-made objects categorized as (C) vehicle at the superordinate level or (D) car at the basic level.

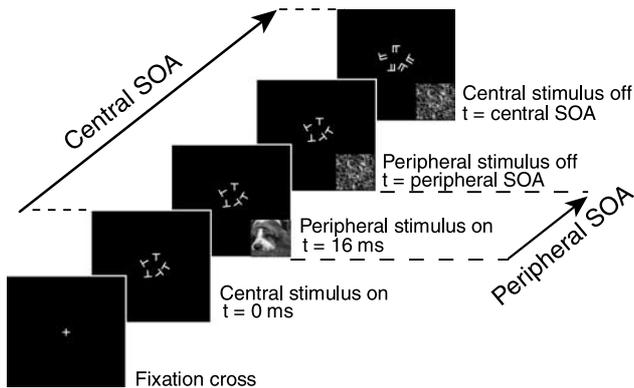


Figure 2. Schematic timeline for one trial in the dual-task paradigm. At the end of a trial, participants were asked to perform the central letter discrimination task (whether all letters were the same or one was different from the others) and/or the peripheral task (e.g., whether a dog was present or not). The display was the same for all conditions and only the instructions differed. The letters and the peripheral stimulus were masked individually. The central SOA and the peripheral SOA indicate the presentation time for the letters (~215 ms) and for the dog/non-dog task (~180 ms), respectively. They were adjusted for each subject and each task.

details). A natural scene was presented 16 ms later in the periphery. The stimulus durations for the central and peripheral stimuli were determined individually for each subject (see below for the procedure). The central and peripheral stimuli were masked with the constraint that the peripheral stimulus was always masked before the central stimulus. In this paper, the term “SOA” refers to the time between the onset of the stimulus and the onset of its mask. Thus, the central SOA was always longer than the peripheral SOA. At the end of the trial, subjects made behavioral reports on the central and/or peripheral tasks depending on the task condition (single or dual task). Participants were asked to be as accurate as possible and no constraint was placed on their reaction times. An auditory tone was provided as feedback for incorrect answers for each task in all conditions.

The central attentionally demanding task: A letter discrimination task

Each trial started with a fixation cross appearing at the center of a black screen for a random duration between 300 and 400 ms before the first stimulus appeared. At 0 ms, five letters, randomly oriented Ls and Ts, were presented in the center of the screen at 5 of 9 possible locations within a radius of 1.2° from the center. All five letters could be the same (5 Ts or 5 Ls) or one of them could be different from the other four (1 L among 4 Ts or 1 T among 4 Ls). Participants were asked to report

whether all letters were the same or not by pressing one of two keys on the keyboard with their left hand. Each letter was individually masked by an F, rotated by an angle corresponding to the T or L it replaced. The central SOA was determined independently for each subject (see experimental procedure Step 1 below) and was the same in both single- and dual-task conditions.

Peripheral task

The peripheral stimulus (natural scene or disk), subtending $5^\circ \times 5^\circ$ of visual angle, was presented 16 ms after the onset of the central stimulus at a point on the edge of an imaginary rectangle $12^\circ \times 10^\circ$ in size and masked at the end of its presentation duration. It was always masked before the central stimulus was masked. Participants had to perform 3 different peripheral tasks in each experimental session: a categorization at the basic level (task A) and two control tasks: a categorization at the superordinate level previously reported not to require focal attention (Task B) and a control disk discrimination task known to be attentionally demanding (Task C; Braun & Julesz, 1998; Fei-Fei et al., 2005; Li et al., 2002; Reddy et al., 2004; VanRullen et al., 2004). Each peripheral task was performed twice in the single-task condition and 3 times in the dual-task condition in separate blocks. Participants reported their answer by pressing one of two keys on the keyboard with their right hand.

A peripheral face–gender discrimination task was used during dual-task training (see Step 1 in Table 1).

Dual task: Experimental procedure for training and testing

In separate blocks of the experiment, participants performed each peripheral task (A, B, C) simultaneously with the central letter task. They were asked to fixate the center of the screen and give priority to the central task such that performance on this task would not be significantly different in the single- and the dual-task conditions. Participants were also required to perform the peripheral task as well as possible.

Performing at a stable level in this dual-task paradigm needs a lot of training. Moreover, the dual-task paradigm is valid only if the SOAs on each task are well defined. Indeed, if the stimulus presentation times are long enough, participants will have ample time to shift their attention from one task to the other in the dual-task condition. Therefore, for each participant and each of the central and peripheral tasks, we determined the SOA at which performance in the single-task condition would be between 75 and 85%. These SOAs were determined in 2 separate training steps (Steps 1 and 2 below) before the final test sessions (Step 3 below). Thus, the experiment consisted of 3 stages (see Table 1).

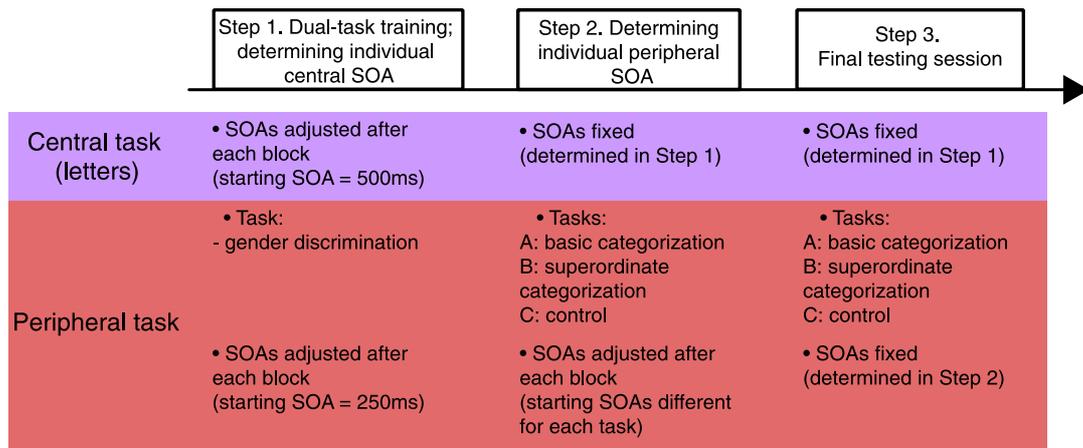


Table 1. Procedure followed by each subject to determine the SOAs required in each task. SOAs were adjusted after each block as follows: They were decreased by 10 ms when participants’ performance in the single-task condition was above 85% and increased by 10 ms when it fell below 75%. Note that the peripheral SOA determined in Step 1 is not used subsequently as it refers to a categorization task (female/male faces) that we used to train subjects on the dual-task paradigm and to determine central SOAs. Peripheral SOAs relevant for the present experiments are individually determined in Step 2.

Step 1: Dual-task training and determining individual central SOA

The first step of the experiment was a training period of approximately 7 h that allowed participants to familiarize themselves with the dual-task paradigm and, more importantly, that allowed the SOAs on the central task to stabilize. Both goals were achieved concurrently by mixing single-task blocks (central or peripheral) with dual-task blocks. A session was composed of 4 blocks of the central letter task in the single-task condition, 4 blocks of a peripheral task in the single-task condition, and 6 blocks of the dual-task condition. During the experiment, the central and peripheral SOAs were decreased by 10 ms each time the mean performance, in a 48-trial block, was above 90%. At the beginning of the training, the letters were displayed for 500 ms (Reddy et al., 2004). To avoid overlearning the main task and stimuli (animals and vehicles), participants performed a face–gender discrimination task in the periphery during this initial training phase (the peripheral SOAs for the natural scenes were determined in Step 2 (see below)). The starting face SOA was 250 ms (Reddy et al., 2004). This stage was completed when the SOAs on the letter task were stable for three consecutive sessions. In the present experiment, the central SOA had to be longer than the peripheral SOA in order to prevent subjects from shifting their attention to the peripheral task after completing the central task (as mentioned above, participants were instructed to prioritize the central task). One subject (GV) performed extremely well on the central letter task resulting in a shorter central SOA than peripheral SOA. Hence, for this subject, we increased the difficulty of the central task by presenting 9 letters instead of 5 so that the central SOA for this

participant was longer than the peripheral one. After this training stage, the central SOAs determined individually for each participant varied from 170 for subject GV up to 260 ms for subject GBJ (mean ≈ 215 ms). These central SOAs were then used in Steps 2 and 3 of the experiment (see below) regardless of the peripheral task.

Step 2: Determining individual peripheral SOAs

In this phase of training, the SOAs for the peripheral tasks of interest were determined for each task and each participant. In particular, as described above, participants performed 3 different peripheral tasks: a basic-level categorization (Task A), a superordinate-level categorization (Task B), and a color pattern discrimination task (Task C, red half on the left or right) known to be attentionally demanding. In this phase, as well as in the data collection phase (Step 3), a session consisted of 19 blocks of 48 trials each: 4 blocks of the central letter task in the single-task condition, 6 blocks of the peripheral tasks in the single-task condition (2 Task A, 2 Task B, 2 Task C), and 9 blocks of the dual-task condition (3 of each task). Block order was randomized. During training Step 2, based on a previous experiment (Li et al., 2002) and pilot data, the starting SOA was set at 170 ms for the basic-level categorization task and at 100 ms for the superordinate-level categorization and color pattern discrimination tasks. SOAs were then decreased by 10 ms when participants’ performance in the single-task condition was above 85% and increased by 10 ms when it fell below 75%. The participants went through to the next stage when their SOAs were stable for three consecutive sessions. On average, training Step 2 took 4 h. In order to avoid any

bias related to image learning, the stimuli seen by a participant during this step were not used in the final testing session.

Step 3: Final testing session

Once training was complete and SOAs defined for each participant and each task (central task and peripheral tasks A, B, and C), data were collected for 10 sessions in Step 3. As described above, a session consisted of 19 blocks of 48 trials each. A session was considered valid if performance on the letter discrimination task was not significantly lower in the dual-task condition than in the single-task condition (paired *t*-test, n.s.). This ensured that participants focused their attention on the central letter task. Over all sessions and participants, only 4 sessions (3 in Experiment I and 1 in Experiment II) out of 102 were rejected by this criterion.

Data analysis

Once normality of data was verified with a Shapiro–Wilcoxon test, a 2-way repeated measures ANOVA and paired *t*-tests were performed to compare single- and dual-task performance, for the superordinate- and basic-level categorization tasks. To summarize and compare participants' performance across different stimulus types, results in the dual-task condition were normalized with respect to the corresponding single-task performance. For each subject, mean performance in the single-task condition is taken as 100% and chance as 50%. Thus, in the dual-task condition,

$$\text{Normalized performance} = 1/2 + 1/2[(P2 - 1/2)/(P1 - 1/2)], \quad (1)$$

where P2 and P1 refer to the mean performance in the dual- and single-task conditions, respectively.

Experiment I: Basic-level categorization of natural stimuli

As described above, attentional requirements were investigated in the performance of three peripheral tasks, all performed in the same experimental session. Participants performed a basic dog/non-dog animal categorization task (Task A) and a superordinate-level categorization task (Task B: animal/non-animal) in order to compare our results with previous ones (Fei-Fei et al., 2005; Li et al., 2002). Furthermore, to verify the efficacy of our protocol, they performed a color pattern discrimination task (Task C), known to be attentionally demanding. Over a period of five days (two 1/2 h sessions per day), participants performed 10 experimental sessions in

Experiment I, in each of which they were tested on all 3 peripheral tasks (A, B, and C). In each of these sessions, they performed a total of 19 blocks of 48 trials: 4 single central task blocks, 6 single peripheral task blocks (2 Task A, 2 Task B, and 2 Task C), and 9 dual-task blocks (3 of each task). Conditions were random and instructions about the condition (single- or dual-task condition) and the peripheral task (Tasks A, B, and C) were given at the beginning of the block.

Task A: Basic-level categorization of natural scenes

Task A was basic-level categorization of natural images. Subjects were asked to report whether the scene contained a dog or another animal. Half the images included at least one dog, and the other half included other animals. A random half of the dog images used in this task was also used as animal images in the superordinate task (Task B). This meant that for this subset of images subjects performed categorizations at both the superordinate and basic levels in separate blocks. The presentation of the peripheral stimulus was followed by a mask (8 different masks were used, composed of a mixture of white noise at different spatial frequencies on which a gray texture was superimposed). Peripheral SOAs for the dog/non-dog task determined for each subject separately (described in experimental procedure Step 2) ranged from 140 for subject GV to 220 ms for subject GBJ (see Table 2).

Task B: Superordinate-level categorization of natural scenes

For Task B, participants performed superordinate-level categorization of natural images and reported whether a scene contained an animal or not. In this experiment, half of the natural scenes contained one or more animals and the other half one or more vehicles. Among the animal stimuli, half the images contained dogs, and the other half had other animals. The same 8 different masks used for the basic-level categorization followed the presentation of the peripheral stimulus. Individual peripheral SOAs (experimental procedure Step 2) for the animal/non-animal task

	Central SOA (ms)	Peripheral SOAs (ms)		
	Letters	Dog	Animal	Disk
MP	220	170	90	70
LR	200	170	100	50
GV	170	140	60	60
GBJ	260	220	110	90
LD	250	205	80	60
RV	200	170	70	70

Table 2. Stimulus presentation time (SOA) used in the final testing session for the 6 participants in Experiment I.

varied from 60 ms for subject GV to 110 ms for subject GBJ (see Table 2).

Task C: Color pattern discrimination

In different blocks, participants also performed a peripheral task known to be attentionally demanding in order to establish that the dual-task paradigm efficiently withdraws attentional resources from the periphery. For this task, the peripheral stimulus was a disk vertically divided with two colors, red and green, in each half. Participants reported whether the red half of the disk was on the left or right side. The presentation of the disk was followed by a mask (6 different masks composed of irregular patches of red and green were used). SOAs for this task were determined individually for each subject (experimental procedure Step 2) and ranged from 50 ms for subject LR to 90 ms for subject GBJ (see Table 2).

Experiment II: Basic-level categorization of man-made stimuli

This experiment was similar to Experiment I except that for the peripheral task subjects performed basic- and superordinate-level categorizations of man-made stimuli (Task A: car/non-car and Task B: vehicle/non-vehicle). They also performed the disk discrimination task as a control (Task C). Out of the 6 participants, 4 had performed Experiment I. At the end of training (Steps 1 and 2), the peripheral SOAs varied from 75 ms for subject LD to 110 ms for subject RC for the superordinate-level categorization task and from 105 ms for subject GV to 175 ms for subject LD for the basic-level categorization task (see Table 3). Participants performed at least 7 test sessions for the final data collection stage.

Experiment III: Effect of stimulus presentation time in the peripheral task

Peripheral SOAs were always longer for basic than for superordinate categorization. In a third experiment, we

	Central SOA (ms)	Peripheral SOAs (ms)		
	Letters	Car	Vehicle	Disk
MP	220	160	70	70
GV	170	105	75	60
GBJ	260	145	90	90
LD	250	175	75	60
RS	200	140	100	120
RC	190	130	110	80

Table 3. Stimulus presentation time (SOA) used in the final testing session for the 6 participants in Experiment II.

determined whether the information available to the subjects when successfully categorizing animals and vehicles at the superordinate level in Experiments I and II could also lead to successful categorization at the basic level. In other words, when subjects can successfully determine whether an image contains an animal or not, can they actually report what kind of animal they have detected? To address this question, in Experiment III, subjects performed a basic-level categorization but with stimulus durations that were limited to those obtained in the superordinate-level task in earlier experiments. Five subjects from Experiment I were tested on biologically relevant stimuli (animal/non-animal and dog/non-dog) and 5 subjects from Experiment II were tested on man-made stimuli (vehicle/non-vehicle and car/non-car). Note that the same experimental design as in Experiments I and II was used (both central and peripheral stimuli were presented), but participants performed only the peripheral categorization tasks in the single-task condition, when attentional resources were fully available (Rousset, Thorpe, & Fabre-Thorpe, 2004).

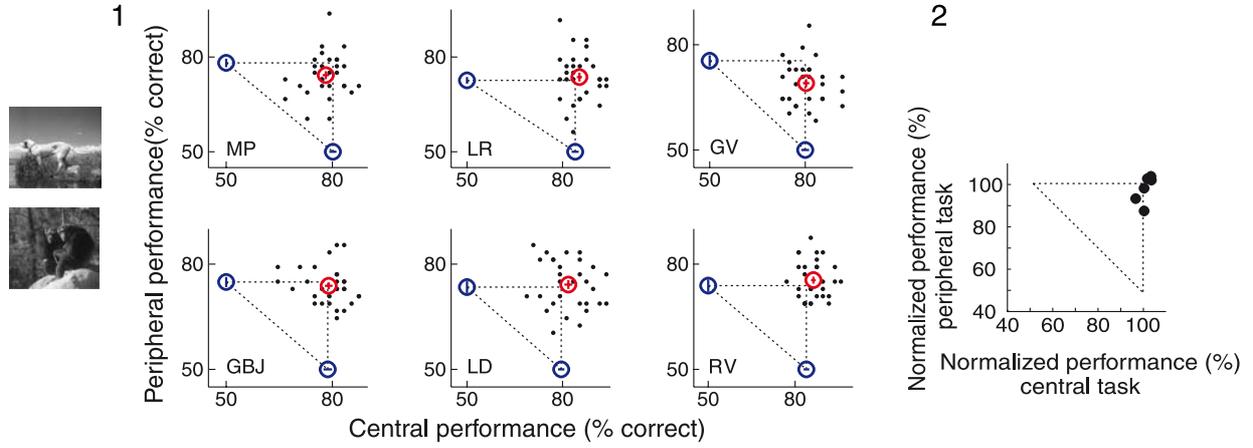
Results

In this study, we used a dual-task paradigm to determine the attentional requirements of various categorization tasks. Participants were required to perform a central attentionally demanding task (letter discrimination) and a peripheral categorization task. Experiment I tested biologically relevant objects (i.e., categorization of animals) in the periphery and Experiment II tested man-made objects (i.e., categorization of vehicles). The central and the peripheral tasks were either performed separately (single-task condition) or simultaneously (dual-task condition). In the dual-task condition, subjects were instructed to prioritize the central letter task such that performance on this task in the single- and dual-task conditions would be equivalent. Indeed, if this criterion was not satisfied in each session, the corresponding session was rejected from all subsequent analyses. The role of attention was measured by comparing participants' performance for the categorization task in the single-task condition (when attentional resources were available) and in the dual-task condition (when spatial attention was focused on the central task).

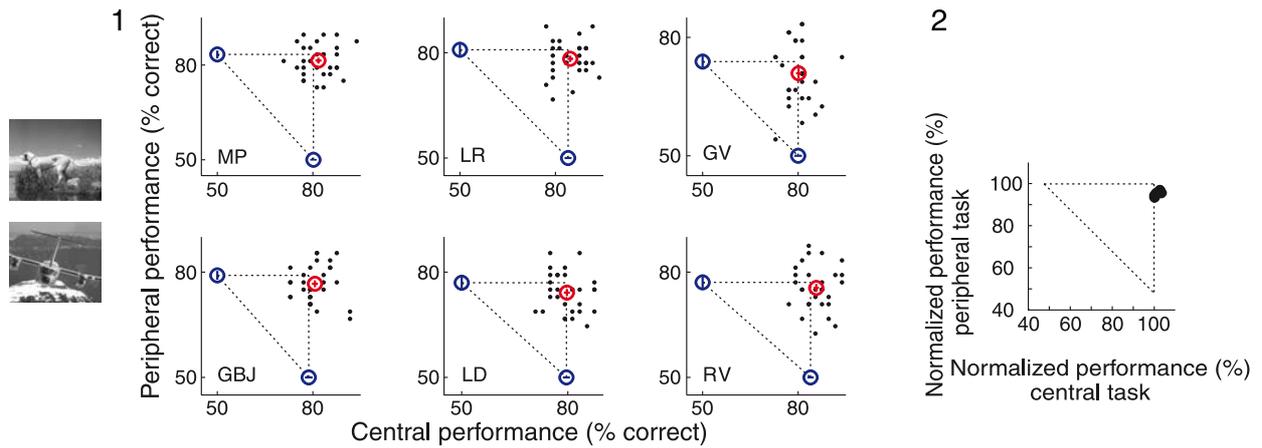
Experiment I: Basic-level categorization of natural stimuli

In the first experiment, we tested whether biologically relevant objects could be categorized at the basic level while attention is engaged on a demanding central letter discrimination task. In particular, we compared the

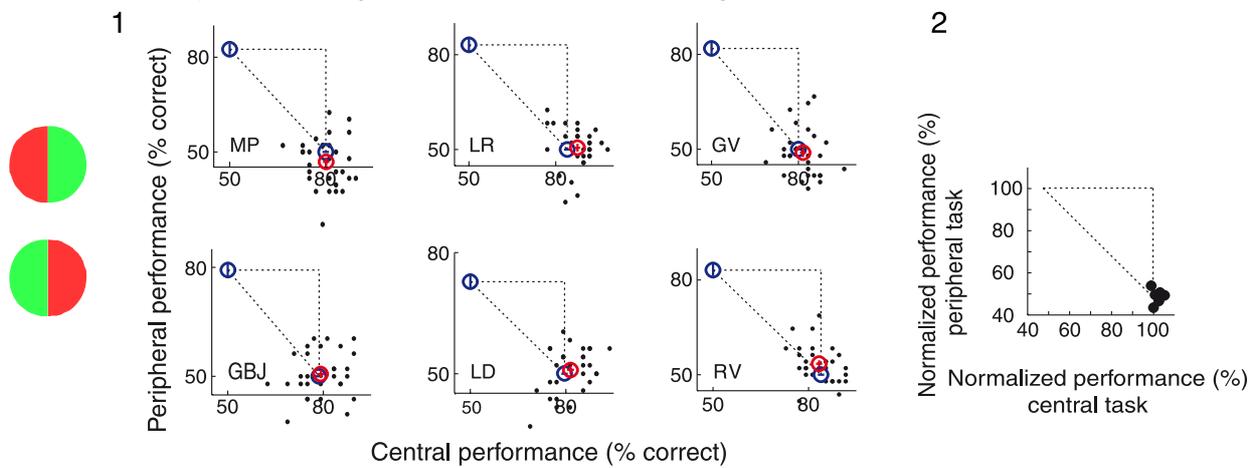
A Basic-level categorization: dog/non-dog animal



B Superordinate-level categorization: animal/non-animal



C Attentionally-demanding task: red half on the left/right



performance of participants when performing a dog/non-dog discrimination task either alone or in the dual-task condition. Single-task performance in the dog/non-dog animal discrimination task was $74.7\% \pm 2.0\%$ and performance on the same task in the dual-task condition was $71.4 \pm 2.0\%$ (Figure 3A1). These performance values were not significantly different from each other (paired t -test, $p = 0.35$). Moreover, individual t -tests for each participant showed that there was no significant difference between the single- and dual-task performance for five of the six participants (paired t -test, $p > 0.05$). This applies for all participants after correction for multiple comparisons (Bonferroni method). To summarize these results, we calculated the normalized performance on the basic-level categorization for each subject in the dual-task condition as a function of their performance in the single-task condition (Figure 3A2, see Methods section). Normalized performance for the group of participants in the dual-task condition was above 85% of the performance in the single-task condition. These results indicated that even though there was a slight decrement in accuracy when attention was not fully available, performance values were still remarkably high and basic-level categorization of natural stimuli could be performed well with little or no attentional resources.

Previous studies have shown that natural scene categorization at the superordinate level (i.e., animal/vehicle) can be performed outside the focus of attention (Li et al., 2002; VanRullen et al., 2004). To compare the performance of our subjects with these previous reports, we also tested them on a similar animal/non-animal task (Figure 3B1). Normalizing these results as described above indicated that participants were able to perform the superordinate-level categorization task in the dual-task condition at above 90% of their performance in the single-task condition (Figure 3B2). This result confirms previous findings that the animal/non-animal categorization task can be performed in the near absence of attention.

We then compared behavioral performance between the two levels of categorization (superordinate and basic) and

the two attentional conditions (single and dual), with a 2×2 repeated measures ANOVA. Performance was slightly impaired in the dual-task condition compared to the single-task condition ($F(1,5) = 7.28$, $p = 0.04$), and performance in the superordinate-level categorization task was superior to that in the basic-level categorization task ($F(1,5) = 9.76$, $p = 0.03$) in both attentional conditions. The interaction between level of categorization and attentional condition was not significant ($F(1,5) = 0.7$, $p = 0.4$), indicating that withdrawing attention similarly affects both levels of categorization. These results demonstrate that categorization of natural stimuli at the basic level can be performed as efficiently as at the superordinate level, when attention is focused on a central attentionally demanding task.

The interpretation of these results is based on the supposition that the central letter discrimination task is effective at engaging participants' attentional resources at the center of the screen, implying that performance on attentionally demanding tasks in the periphery should suffer substantially in the dual-task condition. To verify that this was indeed the case, we included a control condition in which subjects were tested in the periphery on a color pattern discrimination task known to be attentionally demanding (Li et al., 2002; VanRullen et al., 2004). Specifically, participants were asked to discriminate between two bisected colored disks as shown in Figure 3C. Participants received the same amount of training on this task as for the superordinate and basic levels of categorization. Contrary to the results obtained for natural object categorization, a dramatic decrease in performance was observed for the six participants in the disk discrimination task when it was performed in the dual-task condition ($50.0 \pm 2.3\%$) compared to when it was performed in the single-task condition ($80.4 \pm 4.0\%$; paired t -test, $p < 10^{-4}$; Figure 3C1). Indeed, performance in the dual-task condition was not significantly different from chance level (paired t -test, $p = 0.8$). Normalized performance in the dual-task condition for this task was between 44% and 56% of the level of performance obtained

Figure 3. Results of six participants in the dual-task paradigm for biological and artificial stimuli in the periphery. (1) Individual results. The horizontal axis represents performance on the central attentionally demanding letter discrimination task. The vertical axis represents accuracy (A) on the peripheral basic-level categorization task (dog/non-dog animal), (B) on the superordinate-level categorization task (animal/non-animal), and (C) on the color pattern discrimination task. Participants' mean performance is represented by a blue circle in the single-task condition (single central task and single peripheral task) and a red circle in the dual-task condition. Each black point represents participants' performance for a 48-trial block in the dual-task condition. For plotting purposes, we assume that in the single-task condition performance on the other task was at chance (50%). The error bars represent the SEM. For all participants, performance in the dual-task condition was not significantly different from performance in the single-task condition (paired t -test, $p > 0.05$) except GV in the dog/non-dog animal task (A1) and except GBJ in the animal/non-animal task (B1). On the contrary, performance of all participants in the color pattern discrimination task (C1) was dramatically impaired in the dual-task condition compared to the single-task condition (paired t -test, $p < 10^{-5}$). (2) Normalized results. Each circle represents the mean of one participant's performance in the dual-task condition, normalized by his/her performance in the single-task condition. Normalized values are obtained by a linear scaling that maps the average single-task performance to 100%, leaving chance at 50% (see Methods section). These results demonstrate that participants cannot perform an attentionally demanding task when attentional resources are removed from the periphery (C2), but discriminating biologically relevant stimuli at the superordinate (B2) and basic levels (A2) is robust even in the near absence of attention.

when the task was performed alone (Figure 3C2). These results confirm that the attentional requirements of the central letter task lead to a clear decrease in performance in the dual-task condition when the peripheral task also requires attention. In contrast, performance on the basic and superordinate level categorization tasks was far from being at chance (paired t -test, $p < 10^{-4}$) and significantly different from performance obtained on the disk discrimination task in the dual-task condition (paired t -test, $p < 10^{-4}$).

Experiment II: Basic-level categorization of man-made stimuli

It is possible that the results obtained in Experiment I could be explained by evolved neural networks that specifically process biologically pertinent stimuli (New, Cosmides, & Tooby, 2007). Thus, in a second experiment, we tested if other types of stimuli that appeared relatively recently in our environment, and are necessary learned during our lifetime, could also be categorized at the basic level without a need for attentional resources. Specifically, we tested the attentional requirements of a car/non-car vehicle discrimination task using the same dual-task paradigm as in Experiment I.

Mean performance on the basic-level categorization task (car/non-car) was $82.1 \pm 2.8\%$ in the single-task condition and $78.1 \pm 3.4\%$ in the dual-task condition (Figure 4A1). Even though performance was slightly lower in the dual-task condition compared to the single-task condition, it was still well above chance (paired t -test, $p < 10^{-5}$). On the contrary, when these participants were tested on the disk discrimination task used in Experiment I, their performance in the dual-task condition was not different from chance (paired t -test, $p = 0.7$) and well below performance obtained on the car/non-car task in the dual-task condition (paired t -test, $p < 10^{-4}$). These results show that participants were able to categorize man-made stimuli efficiently even when their attention was engaged by an attentionally demanding task. Moreover, normalized performance of each participant for the car/non-car categorization in the dual-task condition was above 90% (Figure 4A2). Thus, basic-level categorization of man-made stimuli is also possible when attentional resources are not fully available.

Again, to compare the performance of our subjects with previous studies reporting that superordinate-level categorization tasks of man-made stimuli can be performed in the near absence of attention (Fei-Fei et al., 2005; Li et al., 2002), we also tested our participants on a vehicle/non-vehicle task (Figure 4B1). Each participant's normalized performance in the dual-task condition was above 85% of their performance in the single-task condition (Figure 4B2), which confirms previous results.

As in Experiment I, a 2×2 repeated measures ANOVA (single-/dual-task condition \times superordinate/basic categorization) was computed. Performance was higher in the single-task condition than in the dual-task condition ($F(1,5) = 30.19$, $p < 0.05$) and there was no effect of the level of categorization ($F(1,5) = 3.31$, $p = 0.13$). Again, the interaction between the main factors was not significant ($F(1,5) = 1.56$, $p = 0.3$) indicating that withdrawing attention similarly affects both levels of categorization. Man-made stimuli seem to suffer slightly more than natural stimuli when they are categorized outside the focus of attention both at the superordinate and basic levels. However, normalized performance in the dual-task condition was still above 85% of the performance obtained in the single-task condition and did not drop to chance as observed for the disk discrimination task. Basic- and superordinate-level categorization tasks of man-made stimuli can be performed efficiently in the near absence of attention.

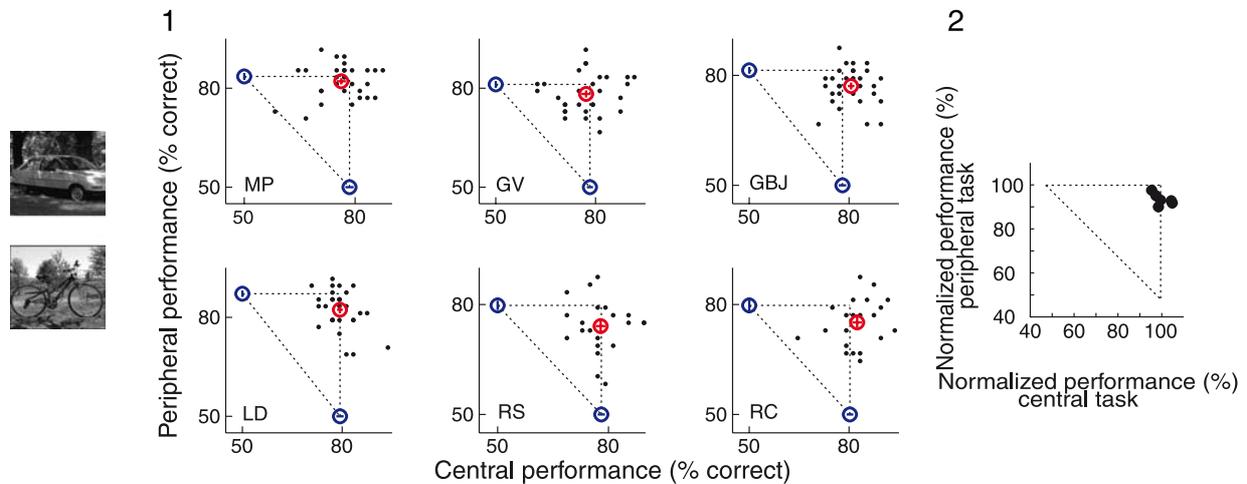
Biologically relevant or not, object categorization at the superordinate level and at the basic level was performed efficiently concurrently with an attentionally demanding task. The summary of the results obtained in Experiments I and II is represented in Figure 5. Mean normalized performance for all natural scene categorization tasks lay above 90%, whereas mean normalized performance for the disk discrimination task was at chance level.

Experiment III: Effect of stimulus presentation time on peripheral task performance

For both the superordinate- and basic-level categorization tasks, we equated task difficulty by adjusting the SOAs such that performance in the single-task condition was approximately 75% (see Methods section). However, to achieve such levels of performance, the basic-level tasks required SOAs that were approximately twice as long as those of the corresponding superordinate-level tasks (~ 180 ms for the dog/non-dog task and ~ 140 ms for the car/non-car compared to ~ 85 ms for the animal/non-animal task and ~ 87 ms for the vehicle/non-vehicle task). This finding suggests that at short exposure times, participants could reliably detect an animal or a vehicle in a scene (superordinate-level category) in the near absence of attention but might be unable to access basic-level details of that scene (Fei-Fei et al., 2007; Greene & Oliva, 2009; Joubert et al., 2007). To address this question, we asked participants to perform basic-level categorization using the same shorter SOA that allowed for successful categorization at the superordinate level in Experiments I and II.

In Experiment III, the same experimental paradigm as in Experiments I and II was used, but participants were asked only to perform the peripheral categorization tasks

A Basic-level categorization: car/non-car vehicle



B Superordinate-level categorization: vehicle/non-vehicle

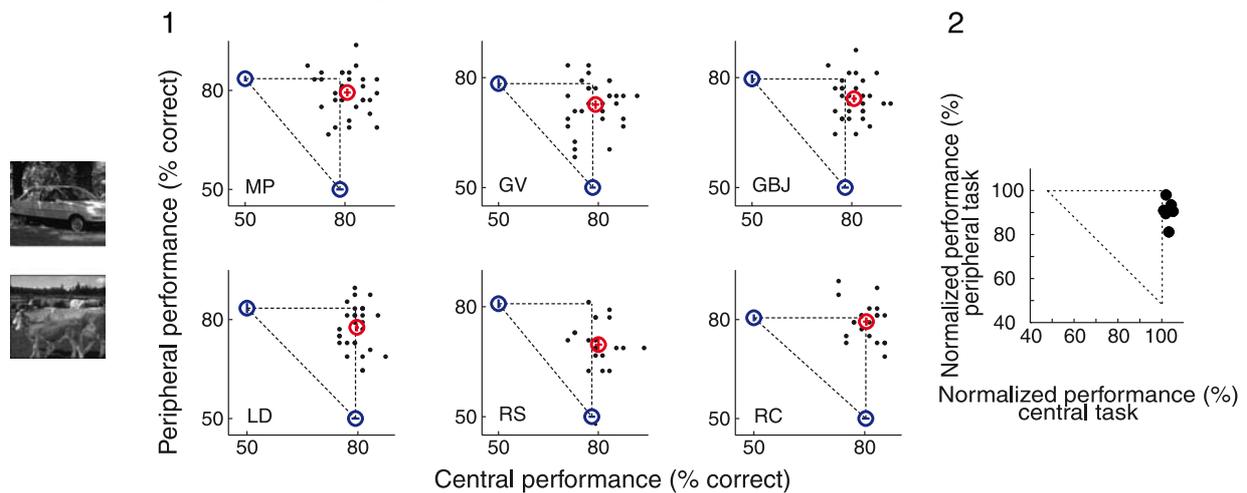


Figure 4. Results of six participants in the dual-task paradigm for man-made object categorization tasks in the periphery. Legend as in Figure 3. Performance in the dual-task paradigm was tested for peripheral categorization at the (A) basic (car/non-car vehicle) and (B) superordinate levels (vehicle/non-vehicle). There was no significant difference in performance between the single-task condition and the dual-task condition (paired t -test, $p > 0.05$) on the basic-level categorization task for MP and GBJ (A1) and for MP, GV, and RC on the superordinate-level categorization (vehicle/non-vehicle) task (B1). The normalized dual-task performance for the basic-level categorization task (A2) and for the superordinate-level categorization (B2) was above 85% of the performance in the single-task condition. However, when tested in the same paradigm in a color pattern discrimination task, participants' performance was at chance level (not shown here). This suggests that although performance was slightly lower in the dual-task condition compared to the single-task condition, participants could perform man-made stimulus categorization tasks at the basic and superordinate levels in the near absence of attention.

in the single-task condition (when attention was fully available). Subjects already tested in Experiment I were tested on the animal/non-animal and dog/non-dog animal tasks, and subjects already tested in Experiment II were tested on the vehicle/non-vehicle and car/non-car vehicle tasks. The critical variable that was manipulated was the SOAs. For both the superordinate- and basic-level categorization tasks, the SOA used in Experiment III

was the one determined for each subject in Experiments I and II for the superordinate categorization tasks (range from 60 to 110 ms for animal/non-animal and from 70 to 110 ms for vehicle/non-vehicle). As expected, performance on the animal/non-animal task was the same in Experiments I and III (paired t -test, $p = 0.4$; Figure 6B), as was performance on the vehicle/non-vehicle task in Experiments II and III (paired t -test, $p = 0.2$; Figure 6D).

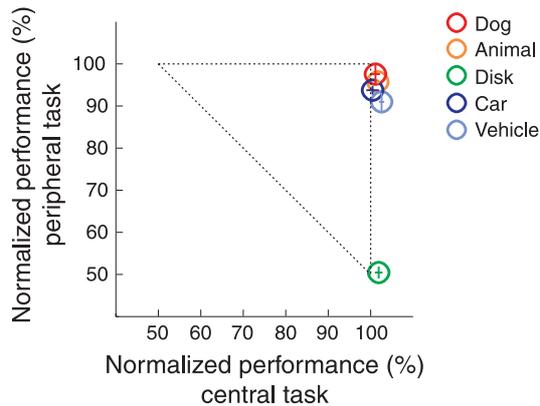


Figure 5. Summary of the results for the peripheral tasks in Experiments I and II in the dual-task condition. Each circle represents the mean of the normalized performance across all subjects for each task. Error bars represent the *SEM*. Natural scene categorization was performed with high levels of accuracy whatever the level of categorization (superordinate or basic) or the nature of the object (biologically relevant: animal, or not: vehicle) when attention was not fully available. In contrast, performance on the disk discrimination task was reduced to the level of chance in the same dual-task paradigm.

On the other hand, a significant drop in performance was observed for basic-level performance. Performance on the dog/non-dog task in Experiment III ($64.5 \pm 4.0\%$) dropped significantly compared to the performance in the same task in Experiment I ($74.8 \pm 2.1\%$, paired *t*-test, $p < 0.01$) when longer SOAs were used (~ 180 ms; Figure 6A). Similarly, performance on the car/non-car task in Experiment III ($75.0 \pm 1.9\%$) dropped significantly compared to performance in Experiment II ($82.3 \pm 3.1\%$; paired *t*-test, $p < 0.05$) when SOAs at around 140 ms were used (Figure 6C). Thus, these results indicate that although participants could reliably report the presence of an animal or a vehicle in a scene with a short stimulus presentation time, they often did not have any further information about what type of animal or vehicle they had just seen. To access information at the basic level, participants needed longer stimulus presentation times, even though they could perform the task in the near absence of attention.

Discussion

The purpose of the present study was to understand the nature of the additional processing observed at both behavioral (Joubert et al., 2007; Mace et al., 2009) and neuronal (Sugase et al., 1999) levels when finer or more detailed information about a stimulus is to be accessed

and, more specifically, to determine if the longer processing time could be explained—at least partly—by a necessary allocation of attentional resources. Previous studies have shown that categorization tasks are feasible at the superordinate level even under conditions when spatial attention is minimally available (Fei-Fei et al., 2005; Li et al., 2002; Rousselet et al., 2002; VanRullen et al., 2004). In the current study, we asked whether finer grained discriminations of these stimuli at the basic level would necessitate the deployment of attention. In other words, although subjects can rapidly detect the presence of an animal in a natural scene (Thorpe, Fize, & Marlot, 1996), without engaging attention (Li et al., 2002; Rousselet et al., 2002), does knowing whether the animal in the scene was (for example) a dog necessarily involve attentional resources? Our results indicate that basic-level categorization of both biologically relevant and man-made

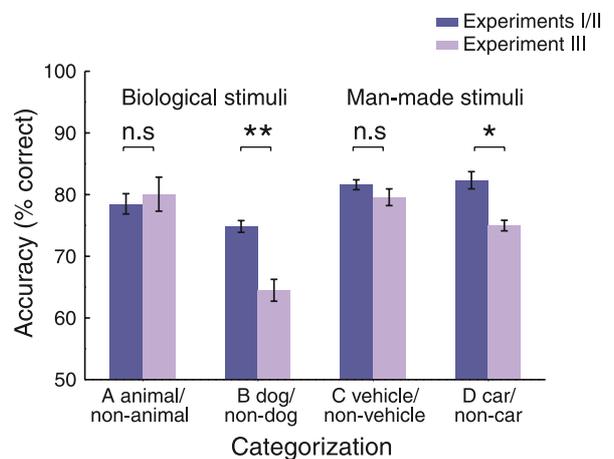


Figure 6. Mean performance on the single peripheral tasks in Experiments I and II (dark color) and in Experiment III (light color). In Experiments I and II, stimulus durations used for the basic-level categorization tasks, dog/non-dog (~ 180 ms) and car/non-car (~ 140 ms), were around twice those used for the corresponding superordinate-level categorization tasks, animal/non-animal (~ 85 ms) and vehicle/non-vehicle (~ 87 ms). With these durations, accuracy was comparable for the two levels of categorization. In Experiment III, we set the stimulus durations for all tasks to those obtained in the corresponding superordinate-level categorization task in Experiment I (~ 85 ms) or II (~ 87 ms) for each participant. Performance on the superordinate-level categorization tasks was comparable between Experiments I and III (A: animal/non-animal task) and between Experiments II and III (C: vehicle/non-vehicle task). However, performance was significantly lower on the basic-level categorization task in Experiment III compared to that obtained with the longer presentation times in Experiment I (B: dog/non-dog task) and in Experiment II (D: car/non-car task). Accessing basic-level information in the periphery required longer stimulus presentation times than accessing superordinate-level information.

objects in natural scenes requires longer presentation times but is feasible even in the absence or near absence of spatial attention.

Presentation time affects rapid stimulus recognition but does not affect the conclusions of this study. In fact, the stimulus SOA required for performing the peripheral task at a given threshold is by no means a predictor of the task's attentional requirements. Indeed, the disk discrimination task can be performed with a short presentation time (~ 75 ms) but cannot be performed in the dual-task condition, unlike natural scene categorization tasks. However, the presentation time of the stimuli may reflect the complexity of the task. Access from coarser to finer object representations (Joubert et al., 2007; Mace et al., 2009; Sugase et al., 1999) requires a longer time for uptake of the relevant visual information, but once this information is obtained, further processing can proceed in the near absence of spatial attention.

Nonetheless, for the natural scene categorization tasks, there was a slight decrement in performance in the dual-task condition compared to the single-task condition (1.3% in the dog/non-dog task and 4% in the car/non-car task). However, small decrements in performance are expected to occur when participants perform two tasks simultaneously and do not necessarily imply a competition for attentional resources. For instance, decrements in performance could also be due to factors such as remembering two different targets or coding and executing two responses simultaneously in the dual-task condition (Duncan, 1980; Pashler, 1994). It is actually quite remarkable to observe such small performance drops in the tasks used here.

It has previously been shown that even when a natural scene was presented very briefly in the periphery (26 ms), observers were able to categorize at the superordinate level fairly accurately (Rousselet et al., 2002, 2004). This seems apparently contradictory to our findings since we used longer presentation times. Such a discrepancy can be explained by the number of locations in which the peripheral stimulus could appear. Indeed, in the above-mentioned studies, there was a drop in accuracy when using four peripheral locations (Rousselet et al., 2004) vs. only two peripheral locations (Rousselet et al., 2002; 80.7 ± 1.1 vs. 90.4 ± 0.6), and in our study, the peripheral stimulus could appear at any point on the edge of an imaginary rectangle (see [Methods](#) section). Moreover, unlike in the present experiments, the stimuli used in the above studies were not masked. We know from backward masking studies that masks effectively interrupt further information uptake from the images (Kahneman, 1968) and that the excitatory neuronal response that occurs just after the disappearance of the target is (Freedman & Miller, 2008). The absence of a mask, therefore, must have allowed for much longer information uptake of the stimulus than what the short presentation duration might imply. By using a mask, we could precisely determine, at least for a peripherally presented stimulus, how much

exposure to the stimulus is needed to accurately categorize a given stimulus at a given level.

Recently, a few studies have obtained contradictory findings about object detection and categorization performance. Initially, Grill-Spector and Kanwisher (2005) proposed that object detection and basic-level categorization may occur at the same early stage of visual processing. However, subsequent studies found results that are inconsistent with this claim (Bowers & Jones, 2008; de la Rosa, Choudhery, & Chatziastros, 2011; Mack, Gauthier, Sadr, & Palmeri, 2008; Mack & Palmeri, 2010). In our study, we show clearly that with the same presentation time, participants could categorize peripheral natural scenes more accurately at the superordinate level than at the basic level (Experiment III). Because images were presented in the periphery in our paradigm, it is possible that any processing differences between category levels, even if small, were amplified, allowing a clear distinction to be manifest, between object detection and categorization. Thus, our results suggest that information about a stimulus accumulates over its presentation duration and, hence, argue in favor of a dissociation between the different levels of object recognition, which is consistent with the finding that superordinate-level categories are processed faster than basic-level categories (Mace et al., 2009).

Natural object categorization (i.e., of biologically pertinent stimuli) could benefit from neuronal populations that have evolved to selectively process these types of stimuli (New et al., 2007). However, our results show that basic-level categorization of natural scenes can be performed in the near absence of attention for not only natural objects but also man-made objects. Given that man-made objects are relatively new in our environment, it is improbable that the ability to categorize visual scenes with little attention is an innate ability (Polk & Farah, 1998). A more likely explanation might be that certain neural networks are predisposed to form representations of stimuli that are often encountered in the environment (Aguirre, Zarahn, & D'Esposito, 1998). Models of visual processing have been developed based on a similar assertion that selective representations for familiar visual stimuli could develop with experience (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007). fMRI studies, showing stimulus-specific activity in the posterior ventro-temporal cortex for categories such as faces and buildings, provide further support for this view (Aguirre et al., 1998; Epstein & Kanwisher, 1998; Haxby et al., 1999; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996). Similarly, category-specific responses have been observed in single neurons in the human medial temporal lobe (Aguirre et al., 1998; Chao, Haxby, & Martin, 1999; Epstein & Kanwisher, 1998; Kreiman, Koch, & Fried, 2000). VanRullen et al. have proposed that such selective representations could underlie visual processing of natural stimuli in the near absence of attention. In contrast, discrimination of unfamiliar stimuli

(e.g., the bisected disks in Experiments I and II, novel stimuli, or low-frequency objects), for which such selective representations would not exist would necessarily involve attentional resources (Fei-Fei et al., 2005; VanRullen, 2009; VanRullen et al., 2004).

Experience and familiarity with stimuli may shape the selectivity of neural networks and enable rapid processing of familiar objects without requiring attention. On the other hand, with longer stimulus durations, the selectivity of neurons in higher order visual areas may increase and activate a deeper representation of the stimulus (Keysers, Xiao, Foldiak, & Perrett, 2001). This may explain why the presentation time required for basic-level categorization is longer than for superordinate-level categorization. However, the activation of these fine representations does not need attentional resources.

Conclusion

In this study, we have shown that when attention is engaged by an attentionally demanding task, basic-level categorization of natural (dog/non-dog animal) and man-made objects (car/non-car vehicle) can still be performed efficiently. This reveals that fine-grained representations of familiar objects can be accessed in the near absence of attention, presumably because they are supported by hardwired, specialized networks (Fei-Fei et al., 2005; VanRullen, 2009; VanRullen et al., 2004). However, this categorization depends on the presentation time of the stimuli (the more detailed the task, the longer the stimulus duration). In other words, we “know” that it is an animal before we “know” that it is a dog, but neither recognition process requires attention.

Acknowledgments

We thank Ramakrishna Chakravarthi for providing helpful comments on the manuscript. This research was supported by the CNRS and the University Paul Sabatier Toulouse III and by a Fyssen Foundation Grant to LR.

Commercial relationships: none.

Corresponding author: Michele Fabre-Thorpe.

Email: mft@cerco.ups-tlse.fr.

Address: CNRS CERCO UMR 5549, Pavillon Baudot, CHU Purpan BP 25202, 31052 Toulouse Cedex, France.

References

- Aguirre, G. K., Zarahn, E., & D’Esposito, M. (1998). An area within human ventral cortex sensitive to “building” stimuli: Evidence and implications. *Neuron*, *21*, 373–383.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*, 285–299.
- Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology*, *61*, 552–557.
- Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4. *Journal of Neuroscience*, *14*, 554–567.
- Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception & Psychophysics*, *60*, 1–23.
- Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics*, *48*, 45–58.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*, 913–919.
- Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron*, *66*, 796–807.
- de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 38–47.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272–300.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1):10, 1–29, <http://www.journalofvision.org/content/7/1/10>, doi:10.1167/7.1.10. [PubMed] [Article]
- Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, *12*, 893–924.
- Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews*, *32*, 311–329.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and

- inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*, 5235–5246.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*, 805–808.
- Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*, 464–472.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition. *Psychological Science*, *16*, 152–160.
- Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*, 189–199.
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13):11, 1–18, <http://www.journalofvision.org/content/8/13/11>, doi:10.1167/8.13.11. [PubMed] [Article]
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*, 3286–3297.
- Julesz, B., & Schumer, R. A. (1981). Early visual perception. *Annual Review of Psychology*, *32*, 575–627.
- Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, *70*, 404–425.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Keysers, C., Xiao, D. K., Foldiak, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, *13*, 90–101.
- Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*, 946–953.
- Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Perception & Psychophysics*, *61*, 1241–1255.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 9596–9601.
- Mace, M. J., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One*, *4*, e5927.
- Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*, 28–35.
- Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1067–1079.
- Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*, 1407–1419.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70–84.
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 16598–16603.
- Pashler, H. (1994). Dual-task interference in simple tasks—Data and theory. *Psychological Bulletin*, *116*, 220–244.
- Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 847–852.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*, 5205–5215.
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2):4, 106–117, <http://www.journalofvision.org/content/4/2/4>, doi:10.1167/4.2.4. [PubMed] [Article]
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of

- surface detail in basic-level object recognition. *Perception*, *33*, 217–236.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*, 629–630.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research*, *44*, 877–894.
- Russell, B., Torralba, A., Murphy, K., & Freeman, W. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*, 157–173.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 6424–6429.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, *17*, 103–119.
- VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience*, *16*, 4–14.
- Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*, 995–1003.

Basic-level object categorization of natural scenes in the near-absence of focal attention

Poncet M.^{1,2,3}, Reddy L.^{1,2}, Fabre-Thorpe M.^{1,2}

¹Université de Toulouse, UPS, Centre de Recherche Cerveau et Cognition, France; ²CNRS, CeReCo, Toulouse, France; ³Université Toulouse II le Mirail, PDPs, France

Introduction

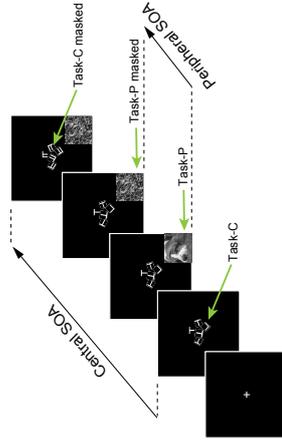
- Superordinate-level categorization (animal/non-animal) can be performed in the near-absence of attention [1].
- Finer categorization (basic-level; dog/non-dog) requires longer reaction times [2].

Can this longer processing time be explained by a necessary allocation of attentional resources?

Method

Dual task (DT) paradigm:

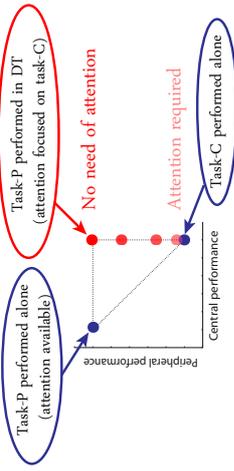
- Central task engaging attention (task-C)
- Categorization task in the periphery (task-P)



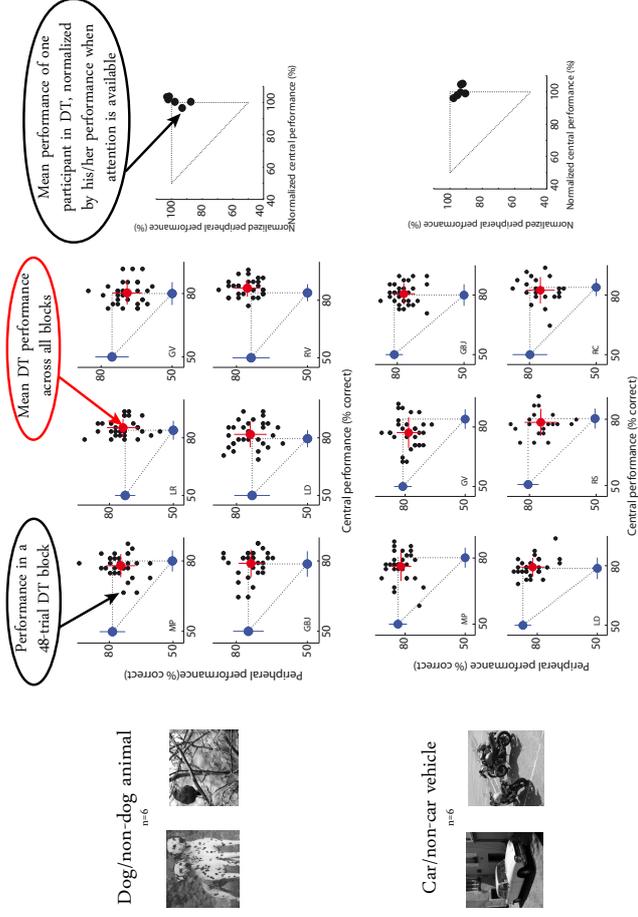
SOAs are adjusted for each participant in each task

Prediction

In DT, priority is given to task-C so that performance on this task should be similar in DT and when performed alone

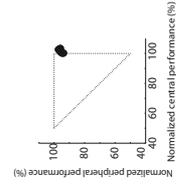


Basic-level categorization: no need of attention

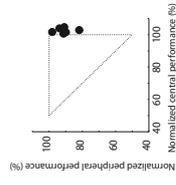


Superordinate-level categorization

Animal/non-animal

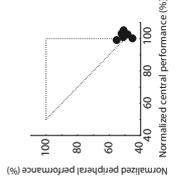


Vehicle/non-vehicle



Attention-demanding task

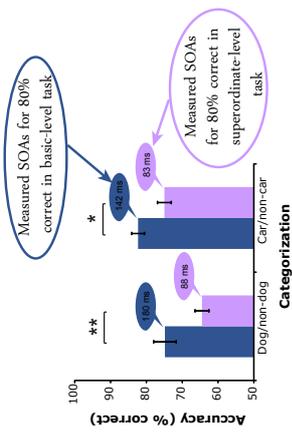
Red half on the left/on the right



Replication of previous results [1]

More information needed

To reach 80% accuracy in task-P performed alone, basic-level categorization requires longer stimulus presentation (SOA_{basic-level} = ~2 x SOA_{superordinate-level})



We "know" that it is an animal before we "know" that it is a dog

Conclusion

Access to basic-level representation can be done:

- Without attention
- Regardless of the category (biological or man-made)
- But requires more time for information uptake

In favor of the hypothesis that familiar representations might be supported by hardwired, specialized networks [3]

References

[1] L.F.F. Verschuuren, R. Koch, C. and Pessoa, P. (2022). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A* 119, 3996-4001.
 [2] Block, M.J., Jodet, O.R., Nepszky, J.L., and Fabre-Thorpe, M. (2009). The time course of visual categorization: you spot the animal faster than the dot. *PLoS One* 4, e6927.
 [3] Verschuuren, R., Reddy, L. and Koch, C. (2024). Visual search and dual tasks reveal two distinct attentional resources. *J Cogn Neurosci* 16, 4-14.

3.3. Discussion

Les résultats de cet article montrent que l'augmentation du temps nécessaire pour catégoriser un objet au niveau basique par rapport au niveau superordonné ne peut pas s'expliquer par une demande en attention différente puisque ces deux tâches de catégorisation peuvent être réalisées en quasi-absence d'attention. Des mécanismes d'intégration de certaines configurations de traits visuels familiers implémentés au niveau neuronal (Rousselet et al., 2004; VanRullen, 2009) pourraient permettre non seulement de réaliser des catégorisations au niveau superordonné mais aussi des catégorisations plus précises au niveau basique en quasi-absence d'attention.

Pour un même temps de présentation des scènes visuelles, et même lorsque les tâches étaient réalisées seules (pas en condition de double tâche), les participants faisaient plus d'erreurs pour des catégorisations basiques que superordonnées. La différence entre les deux niveaux de catégorisation était plus grande pour les objets naturels que pour les objets artificiels. Ces observations confirment nos précédents résultats (article 1) montrant un avantage plus important de la catégorisation superordonnée par rapport à la catégorisation basique pour les objets naturels comparés aux objets artificiels.

3.3.1. Quel type de représentation est créé sans attention ?

Les catégorisations de stimuli complexes peuvent être réalisées en quasi-absence d'attention. Une des critiques adressées à l'étude de F. F. Li et al. (2002) est que les processus de traitement à la base de ces résultats ne permettraient pas d'accéder à une représentation détaillée de la scène visuelle. Dans l'article 2, il semble peu probable (même si ce n'est pas impossible) que seul le « gist » d'une image ou les statistiques de l'image comme suggéré dans le modèle de Wolfe, Võ, et al. (2011) puissent être à l'origine des bonnes performances en double tâche pour des catégorisations telles que chien vs. autre animal ou voiture vs. autre véhicule étant donné que la différence de « gist » est faible entre des objets partageant la même catégorie superordonnée.

Par contre, la détection d'un chien ou d'une voiture pourrait être basée sur un ou plusieurs traits caractéristiques de ces catégories. Evans & Treisman (2005) ont ainsi proposé que les traits ou parties

caractéristiques des objets d'une catégorie cible pourraient être perçus grâce à des traitements en parallèle mais que ces parties ne seraient pas intégrées spatialement. Dans leur étude utilisant un paradigme de RSVP, Evans & Treisman, (2005) ont montré que lorsque les participants avaient correctement identifié la présence d'un animal (73% des essais), ils ne le localisaient correctement dans un des quatre quadrants de l'image que dans seulement 53% des cas (niveau de chance à 33%). En d'autres termes, nous serions capables de déterminer si un animal est présent dans une scène en quasi-absence d'attention puisque nous pourrions détecter certaines caractéristiques propres aux animaux (e.g. un bec, des yeux, de la fourrure) mais ces caractéristiques ne seraient pas intégrées en un percept unique « animal ». En fait, ces caractéristiques intermédiaires sont importantes pour la catégorisation animal/non-animal. L'absence d'yeux, d'une bouche ou de membres dans l'image diminuent significativement les performances de catégorisation en termes de précision et de temps de réaction (Delorme, Richard, & Fabre-Thorpe, 2010).

Cette hypothèse n'est pas contradictoire avec l'idée que certaines configurations spatiales de traits visuels familiers pourraient être intégrées en quasi-absence d'attention (e.g. deux yeux et une bouche). Il n'y a peut-être pas d'intégration des différentes parties d'un objet ensemble mais un traitement automatique de certaines configurations (donc un certain type d'intégration de traits visuels) permettant la catégorisation d'objets au niveau superordonnée et basique en quasi-absence d'attention.

3.3.2. Que veut dire en quasi-absence d'attention ?

Récemment, les conclusions d'une étude ont suggéré que les études en double tâche telle que celle de F. F. Li et al. (2002) échouaient à engager toute l'attention dans la tâche centrale principale (M. A. Cohen, Alvarez, & Nakayama, 2011). Dans cet article, les auteurs montraient dans trois expériences de double tâche (différentes de celle de F. F. Li et al. 2002) qu'au contraire de ce qui avait été observé, le traitement de scènes naturelles nécessitait de l'attention. Ils affirment aussi que « *Given that we followed previous studies in using a task that required the detection and classification of animals and vehicles within a scene, the difference in results between the current study and previous studies (F. F. Li et al., 2002; Rousselet et al., 2002) cannot be attributed to a difference in the task used to assess scene perception* ». Cependant cette affirmation est incorrecte au moins sur deux points :

- Premièrement, les auteurs n'utilisaient pas une tâche de catégorisation animal/non-animal mais les participants avaient le choix entre quatre réponses : (1) la scène contient un animal, (2) la scène contient un véhicule, (3) elle ne contient pas d'animal ni de véhicule, (4) aucune scène n'a été présentée (expérience 3). Les participants répondent en quelque sorte à trois questions au lieu d'une : est-ce qu'une scène a été présentée, si oui contenait-elle un objet, si oui lequel était-ce ? Dans l'expérience de F. F. Li et al., (2002), une question unique était posée : la scène contenait-elle un animal ou pas. En fait, comme les auteurs l'ont remarqué, les performances en double tâche étaient bien meilleures avec un autre type de question (catégorisation de scènes en montagne/plage/autoroute/bâtiment/scène d'intérieur) que dans la catégorisation animal à quatre choix que je viens de décrire.
- Deuxièmement, dans leur expérience les participants n'étaient pas entraînés au paradigme de double tâche (au contraire de F. F. Li et al., 2002). Il est fort possible que sans cet entraînement permettant d'améliorer la coordination et l'automatisation des réponses motrices pour les deux tâches demandées (et qui peut être fait sur une autre tâche comme c'est le cas dans l'article 2), les performances soient diminuées par rapport à la condition où la même tâche est réalisée seule. Il semble difficile ensuite de conclure qu'étant donné que les performances diminuent pour la condition en double tâche le traitement des scènes naturelles demande de l'attention.

Les auteurs ont surtout orienté leurs conclusions sur le fait que les résultats obtenus en double tâche ne pouvaient pas être un exemple de « conscience sans attention ». Cette conclusion semble tout à fait correcte. Il est en effet difficile de conclure avec un paradigme de double tâche que les stimuli périphériques étaient catégorisés en absence totale d'attention. Malgré de multiples précautions prises dans le protocole pour qu'un minimum d'attention soit disponible en périphérie (la difficulté de la tâche centrale et de la tâche périphérique sont ajustées pour chaque participant), il est possible que quelques ressources attentionnelles puissent quand même être allouées en périphérie. D'ailleurs, les études utilisant une double tâche concluent que le traitement du stimulus se fait en *quasi-absence* d'attention et non pas en *absence* d'attention. L'idée importante de ces travaux (dont l'article 2) est que quel que soit le niveau d'attention dédié au traitement du stimulus périphérique (élevé vs. très faible), les participants sont capables de catégoriser les images avec les mêmes performances.



Chapitre 4
Mémoire



4. Encodage à long terme pendant la catégorisation d'objets

Dans les deux chapitres précédents, j'ai montré que la catégorisation de scènes naturelles complexes est très rapide (environ 365 ms de temps de réponse médian en go/no-go pour la catégorisation animal/non-animal) et peut être réalisée en quasi-absence d'attention. Le temps de présentation des images (20 ms ou 500 ms) n'avait pas d'influence sur les performances ce qui montre qu'un niveau de traitement approfondi n'était pas nécessaire pour réaliser cette tâche. Dans ce cas, on peut se demander : jusqu'à quel niveau traitons-nous les images présentées 20 ms dans une tâche de catégorisation visuelle rapide ?

4.1. Niveau de traitement des images lors de catégorisations visuelles rapides

Tous les participants testés dans une expérience de catégorisation visuelle disent avoir répondu de façon automatique, sans réfléchir. Ce type de tâche est vraiment facile à réaliser, le nombre de réponses correctes s'élève à 95% (pour catégoriser des animaux). Par ailleurs les résultats de Crouzet et al. (2010) suggèrent que la catégorisation de stimuli complexes reposerait sur des mécanismes automatiques qu'il serait difficile de contrôler. Dans leur paradigme de choix saccadique (deux images sont présentées de chaque côté d'un point de fixation), les participants avaient beaucoup de mal à saccader vers des véhicules lorsque les distracteurs étaient des visages. Par contre, ils ne faisaient que très peu d'erreurs pour saccader vers des visages lorsque les distracteurs étaient des véhicules. Les mécanismes « top-down » ne joueraient donc qu'un rôle mineur dans ce type de tâche et la catégorisation d'images complexes pourraient être soutenue par des mécanismes très automatisés. Par ailleurs, les traitements requis pour des tâches de catégorisation ne nécessiteraient pas une analyse très détaillée des stimuli. En fait la catégorisation animal/non-animal pourrait être réalisée avant IT (Girard et al., 2008; Kirchner & Thorpe, 2006) rendant les traitements ultérieurs de la voie ventrale non nécessaires.

Pourtant, cela ne signifie pas qu'aucun traitement cérébral n'est effectué après la prise de décision requise par la tâche. Dans une étude récente, Cauchoix, Barragan-Jason, Serre, & Barbeau, (2014) ont proposé que le traitement cognitif d'une image se ferait en trois phases. Dans leur expérience, ils ont utilisé une tâche de catégorisation visuelle rapide manuelle (visage humain vs. animal) pour des stimuli présentés pendant 100 ms. Les auteurs ont utilisé des analyses MVPA appliquées au signal EEG et montré que l'information de la catégorie d'un objet pourrait être présente très tôt, dès 95 ms après l'apparition du stimulus. Les performances du classifieur étaient basées principalement sur l'activité neuronale des aires visuelles postérieures et étaient corrélées avec les propriétés de bas niveau des images vers 75 ms pour le contraste et vers 115 ms pour la taille du visage. Selon les auteurs, cette phase première phase de traitement pourrait refléter des processus de bas niveau implémentés par une première vague feed-forward de V1 jusqu'aux aires occipito-temporales. Une deuxième phase débuterait vers 125 ms après l'apparition du stimulus. Les performances du classifieur étaient alors stables et étaient corrélées avec les réponses comportementales (temps de réponse manuel). Cette période pourrait refléter la construction d'une représentation interne du stimulus, indépendante des

propriétés de bas niveau de l'image. Les performances du classifieur augmentaient encore ensuite vers 185 ms après la présentation du stimulus. Cette troisième étape pourrait indiquer l'implication d'un réseau distribué des aires cérébrales (les performances du classifieur étaient basées sur l'activité cérébrale provenant de toutes les électrodes) et être associée à un accès conscient à la représentation du stimulus.

Des traitements ultérieurs à 185 ms, c'est-à-dire après que les réponses neuronales soient corrélées avec les temps de réponses, sont donc réalisés. Pour autant, est-ce que les images flashées seulement 20 ms dans des tâches de catégorisation rapide peuvent être traitées par l'hippocampe et stockées en mémoire à long terme ? Les résultats de Kreiman et al., (2000) montrent que vers 200-240 ms après la présentation d'un objet visuel, les neurones du cortex temporal médian montrent une activité sélective à la catégorie des objets. Ces neurones sont situés dans des structures impliquées dans les processus de mémorisation (amygdale, cortex entorhinal, hippocampe) et pourraient donc permettre la mémorisation des images. Cependant, ces activités neuronales ont été observées pour des images présentées pendant 1 seconde. Par ailleurs, les performances de reconnaissance d'images présentées dans un paradigme de RSVP diminuent avec l'augmentation du temps entre la présentation de l'image et le test de reconnaissance (Potter, Staub, Rado, & O'Connor, 2002). Les représentations créées pour des images flashées ne semblent donc pas être stockées en mémoire à long terme mais seulement en mémoire de travail ou à moyen terme. En fait, Melcher (2001) a montré que si la répétition d'un stimulus améliorait les performances de reconnaissance, ce n'était pas le cas lorsque cette répétition s'effectuait sur plusieurs jours.

Dans l'article 3, nous avons demandé à des participants de reconnaître des images qu'ils avaient vues précédemment flashées 20 ms. Nous avons testé deux groupes de participants : un groupe avait vu les stimuli seulement une fois et l'autre groupe les avaient vus une fois par jour pendant trois semaines dans une tâche de catégorisation animal/non-animal. Je présente brièvement le protocole et les résultats de l'entraînement sur la tâche de catégorisation animal/non-animal pour le deuxième groupe (les résultats ont été publiés dans Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001) avant de présenter plus en détails l'étude 3.

4.1.1. Entraînement sur le même set d'images familières

Les temps de réponses en catégorisation visuelle animal/non-animal sont très rapides. Fabre-Thorpe et al., (2001) ont voulu voir s'il était possible d'accélérer ces réponses en utilisant des images avec lesquelles les participants seraient très familiers. Pour cela, ils ont demandé à 14 participants de réaliser une tâche de catégorisation en go/no-go pendant 14 jours (distribués sur trois semaines) sur le même set de 200 images. A la fin de cette phase d'entraînement, les participants étaient testés dans une nouvelle tâche de go/no-go incluant des images familières (celles qu'ils avaient vues pendant trois semaines) et des images nouvelles. La distribution des temps de réaction pour les deux types d'images montrent que seules les latences des réponses les plus lentes ont été raccourcies (Figure 23 à gauche). De plus, même si les participants manquaient moins souvent la présence d'un animal dans une image familière que dans une image nouvelle (1.2% vs. 6.3% de « miss »), la différence d' entre les deux types d'images n'était visible qu'à partir des réponses de plus de 400 ms (Figure 23 à droite). L'entraînement ou le niveau de familiarité avec un set d'images donné influence peu les performances de catégorisation animal/non-animal.

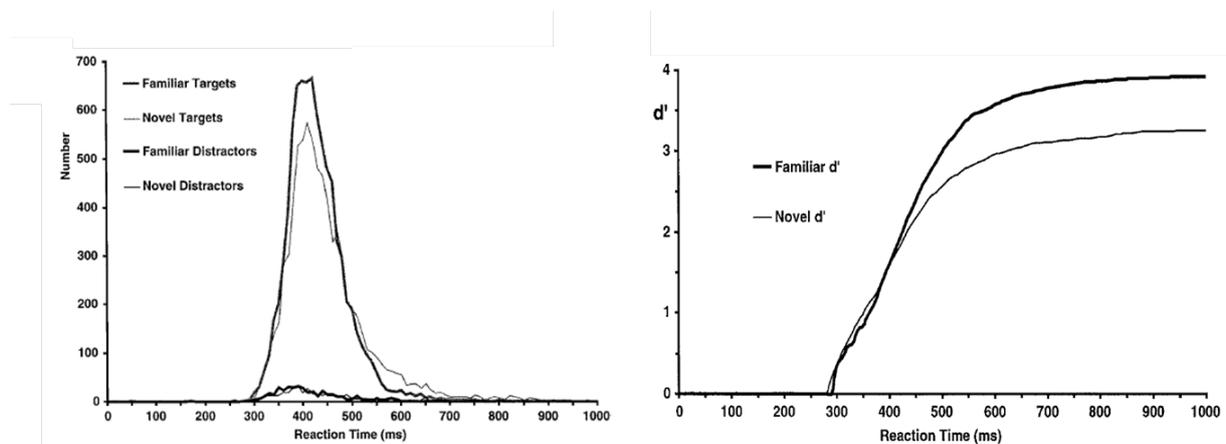


Figure 23. Distribution (par fenêtre de 10 ms) des temps de réaction (à gauche) et représentation de la précision (d') en fonction du temps de réaction (à droite) pour les images familières (en noir) et nouvelles (en gris). (Fabre-Thorpe et al., 2001)

Si on s'intéresse à l'activité enregistrée en EEG lors de la tâche de catégorisation réalisée après l'entraînement, les ERPs pour les stimuli familiers avaient en général une amplitude moins élevée que

ceux des images nouvelles (Figure 24A). De plus, il semble que l'amplitude de la différentielle entre les images d'animaux (cibles) et de non-animaux (distracteurs) soit plus importante pour les images familières à partir de 200 ms après l'apparition du stimulus (Figure 24B). Une trace mnésique pourrait donc avoir été créée pour les images familières (ou au moins certaines) au niveau neuronal mais est-ce suffisant pour reconnaître explicitement ces images parmi des images nouvelles ?

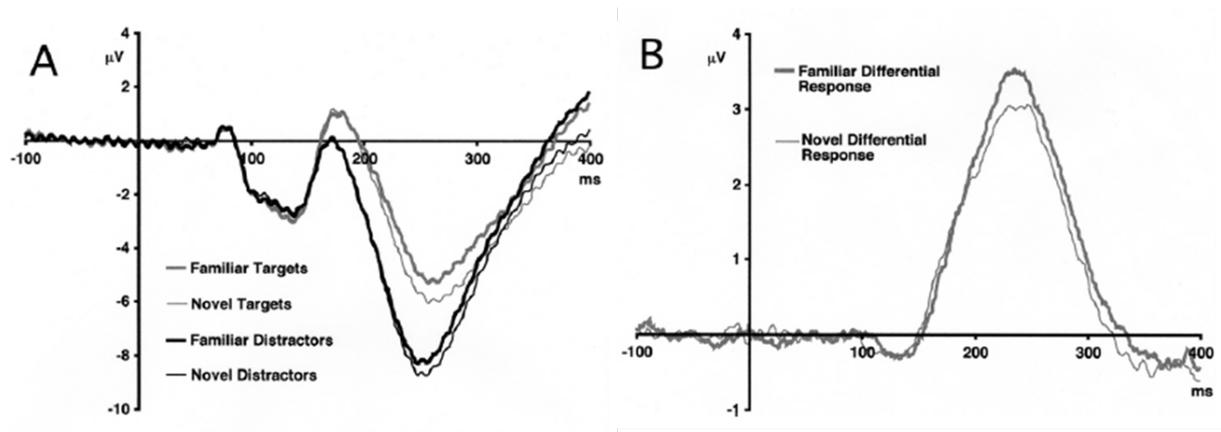


Figure 24. A. ERP de l'électrode Fz moyenné sur les essais corrects de 14 participants pour les images *familières* cibles (animal) et distracteurs (non-animal), ainsi que pour les images *nouvelles* cibles et distracteurs. B. Différence entre les ERP des cibles et des distracteurs pour les essais corrects suivant si l'image était familière ou nouvelle. (Fabre-Thorpe et al., 2001)

La question soulevée par cette expérience est : est-ce que les participants seraient capables de reconnaître le set d'images qu'ils ont vu pendant trois semaines ? En effet, les images n'ont été présentées que pendant 20 ms et on ne demandait pas aux participants de mémoriser les images. De plus, même si les scènes présentées étaient très diverses, elles sont aussi très « commune » (ce sont des scènes naturelles de notre environnement) et parfois très similaires conceptuellement entre elles (en particulier plusieurs images d'un même animal, par exemple un lion, ont été utilisées).

4.1.2. Présentation de l'article 3

Une semaine après avoir été entraînés et testés dans l'expérience décrite dans le chapitre précédent (4.1.1), 13 des 14 participants sont revenus au laboratoire sans avoir été prévenus de la tâche qui allait leur être demandée. Les images étaient présentées pendant 20 ms et les participants devaient donner leur réponse en moins d'une seconde : lever le doigt si l'image était familière et le garder appuyé si l'image n'était pas familière. Les 200 images familières étaient répétées trois fois au cours de l'expérience et mélangées avec 600 images que les participants n'avaient jamais vues.

L'objectif principal de cette expérience était de savoir si les participants entraînés pendant trois semaines sur 200 images en catégorisation rapide animal/non-animal seraient capables de reconnaître ces images parmi d'autres images naturelles complexes. Nous avons aussi testé dans le même paradigme un groupe de 10 participants ayant vu les images une seule fois 20 ms quelques minutes auparavant (groupe non-entraîné). L'ajout de ce groupe nous permet de comparer les performances dans la tâche familier/non-familier pour des participants entraînés ou non-entraînés et ainsi de voir un possible effet de la mémoire à long terme sur la reconnaissance des images. Des analyses supplémentaires ont aussi été réalisées afin de déterminer un effet d'apprentissage au cours de l'expérience avec la répétition (3 fois) des images familières, la différence entre les images animal et non-animal, mais aussi les corrélats neuronaux (signal EEG) des résultats comportementaux.

4.2. Article 3: Briefly flashed scenes can be stored in long-term memory²

² Authors: Arnaud Delorme, Michele Fabre-Thorpe, Marlene Poncet (alphabetic order). MFT designed the study; AD, MFT, MP ran participants; MP analysed the data; MFT and MP interpreted the data; MP wrote the first draft; AD and MFT gave comments on the manuscript.

Briefly flashed scenes can be stored in long-term memory

Abstract

The capacity of human memory is impressive. Previous reports have shown that when asked to memorize images, participants can recognize several thousands of visual objects in great detail even with a single viewing of a few seconds per image. In this experiment, we tested recognition performance for natural scenes that participants saw for 20ms only once (untrained group) or 22 times over many days (trained group) in an unrelated task. 400 images (200 previously viewed and 200 novel images) were flashed one at a time and participants were asked to lift their finger from a pad whenever they thought they had already seen the image (go/no-go paradigm). Compared to previous reports of excellent recognition performance with only single presentations of a few seconds, untrained participants were able to recognize only 59% of the 200 images they had seen a few minutes before. On the other hand, trained participants, who had processed the flashed images (20ms each) several times, could correctly recognize 87% of them. EEG recordings confirmed these behavioral results. As early as 230ms after stimulus onset, a significant ERP difference between familiar and new images was observed for the trained but not for the untrained group. Consistent with previous literature, for both the trained and the untrained group, we also found an automatic processing of visual object categories, demonstrated by an ERP difference between animal and non-animal images starting around 150ms after stimulus onset. These results show that briefly flashed unmasked scenes can be incidentally stored in long-term memory when repeated but a single presentation might not be sufficient.

Introduction

Humans can remember thousands of pictures after a single exposure of 5 sec (Shepard, 1967; Standing, Conezio, & Haber, 1970; Standing, 1973). More impressively, recent studies have shown that the representations stored in memory are very precise (Brady, Konkle, Alvarez, & Oliva, 2008; Hollingworth & Henderson, 2002; Vogt & Magnussen, 2007). For example, in the study of Brady et al., (2008), after viewing 2500 pictures of objects, participants were shown two images and asked which of the two they had seen (two-alternative forced-choice, 2AFC). The new picture could either contain an object from a different category, a new exemplar from the same category, or the same object but in a different state. Impressively, participants successfully discriminated between the previously seen object and the new one at 87% accuracy in the state condition. These results demonstrate that humans have a massive and detailed memory capacity for storing pictures.

In these studies, presentation time of the stimuli was very long (from 3 to 20 sec). However, it is well known that recognition memory suffers when stimulus duration decreases (Potter, 2012; Shaffer & Shiffrin, 1972). For example, Potter (1976) showed that it is possible to remember 80% of masked pictures presented for 120ms but only 50% of them if they were presented for only 50ms. Other studies argued that scenes would first be represented as a general layout or gist and details about specific objects would be added on subsequent fixations (Melcher & Kowler, 2001; Melcher, 2001; Tatler, Gilchrist, & Rusted, 2003). Generally, the amount of detail remembered about a stimulus increases linearly with more time to encode them (Brady, Konkle, Oliva, & Alvarez, 2009; Melcher, 2006). Therefore, long enough stimulus duration seems necessary to encode and recall complex visual scenes in detail.

Nonetheless, the repetition of briefly presented pictures improves memory performance. In a study by Melcher, (2001), participants were presented with scenes including 12 unrelated objects either in one viewing or as repeated brief views (from 0.25 to 2 s) for a cumulative viewing duration of 1 to 4 sec in both conditions. The number of items recalled after 4 s of a continuous presentation was the same as the number of items recalled after 4 trials of 1 s even though each presentation was separated by other stimuli. These results, as well as those of Martini & Maljkovic, (2009) using a rapid-serial-visual-presentation (RSVP) paradigm, are consistent with the Total Time Hypothesis proposing that a fixed amount of time is required to learn a fixed amount of information regardless of how this time is distributed (Bugelski, 1962; Cooper & Pantle, 1967).

The improvement of memory performance due to repetition was tested only after a few minutes and might not involve long-term memory but a “medium term” memory (as referred to by Melcher, 2001). Indeed, the accumulation of memory with re-test trials did not extend across separate days (Melcher & Kowler, 2001; Melcher, 2001). Thus traditional long-term memory was most probably not involved in the increase of performance with repetition. Furthermore, recent results have shown that the memory tested after few minutes in RSVP tasks declines quickly with (a) increased time between the stimulus presentation and test and (b) increased number of intervening test pictures (Endress & Potter, 2012; Potter, Staub, Rado, & O’Connor, 2002). Therefore, even though repetition might improve recognition performance on a short timescale, repeated brief presentations might not be sufficient to encode images in long-term memory.

The first aim of our study was to test visual recognition memory for complex unmasked natural scenes flashed for only 20ms. To further test the effect of repetition on memory performance, one group of participants saw the to-be-recognized images only once and a second group saw the same image set over several days for a cumulative time of 440ms. Such short stimulus presentation time has never been used in long-term memory recognition. This reduces the stimulus energy and prevents the possibility of making saccades to encode details about the stimuli.

We challenged memory performance on two other aspects apart from the use of extremely short stimulus presentation time. First, participants were not explicitly asked to memorize the set of images

and were unaware that they would be tested on a recognition task. Second, instead of the typical 2AFC recognition task, we tested our participants using a go/no-go paradigm. Participants had to decide very rapidly (within 1 s) for each image whether they had seen it earlier (familiar/non-familiar task) without any help from a distractor image (as could be the case in a 2AFC). Long-term memory experiments do not usually require participants to answer in a limited time window. For example, in Brady et al., (2008) participants' reaction times were above 2 s.

To strengthen our results, we investigated a possible correlation between participants' performance in the recognition memory task and event-related-potential (ERP) measures. Since the study of Rugg et al., (1998) an entire branch of research has studied the mechanisms of recognition memory for words and pictures using ERPs. The main goal of these studies was to differentiate familiarity (unspecific awareness that an item has been encountered) from recollection processes (conscious retrieval of both item and contextual details). In general these studies compared ERPs between previously seen and novel images. They reported two distinct components of this old/new effect: an early (300-500 ms) mid-frontal component related to familiarity based recognition and a late (500-700 ms) parietal effect related to recollection (e.g. Curran & Cleary, 2003; Rugg & Curran, 2007; Zimmer & Ecker, 2010; but see also Voss & Federmeier, 2011 for another interpretation of the familiarity effect).

To summarize, we tested along with EEG recordings whether participants could perform a memory recognition task (familiar/non-familiar) on complex images that were flashed only 20ms in an incidental task. Our study challenges memory performance on several aspects: incidental learning, very short stimulus presentation time, and fast recognition.

Method

Participants

The trained group included 13 participants. EEG was recorded in the recognition memory task for only 10 of them. The untrained group included 10 participants (4 women, 3 left-handed, mean age 26 ± 1) all of them tested along with EEG recordings. All participants had normal or corrected to normal acuity and provided written informed consent.

Stimuli

The 2000 images used in the experiment were colored natural images from the Corel database. Half of the images included animals (mammals, birds, insects, etc.), the other half was composed of landscape, fruits, buildings, etc. Images were very diverse but at the same time very similar: different exemplars of the same basic category (e.g. lion) were present in familiar and new images (Figure 1). Therefore, having a vague sense of the presence of a lion in the picture would not be sufficient to distinguish this image from another new image including a lion.

Trained participants were separated into seven pairs. For each pair, a subset of 200 images, half animals and half non-animals, was randomly picked to be the familiar set and all 200 images were different from the 200 familiar images of another pair of participants. In the recognition task, the 200 images were repeated 3 times (see procedure below) while the new images (600) were never repeated and were the same for all participants. The same set of familiar and new images presented to an untrained subject corresponded to the ones presented to a trained subject for whom EEG was recorded. Therefore, the image set was the same for the untrained and the trained participants for whom EEG was recorded.



Figure 1. Examples of animal (top row) and non-animal (bottom row) images used in the experiment.

Procedure

The experiment was composed of 3 stages: a training phase, a study phase and a test phase (Figure 2). The training phase, only performed by the trained group, consisted of an animal/non-animal categorization task on the familiar image set repeated 5 days per week over 3 weeks (for details about each phase, see below). One week after the last training session, both trained and untrained participants saw the 200 familiar images set (study phase). Participants were then asked few minutes later to perform an unexpected recognition test on these images (test stage). EEG was recorded during this last task.

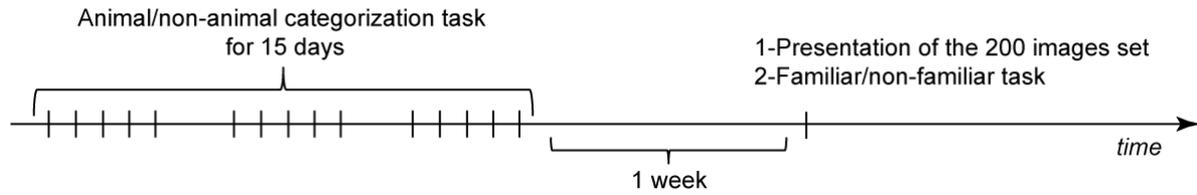


Figure 2. Procedure of the experiment. The trained group performed a categorization task on the familiar image set for 15 days. After a week both trained and untrained participants saw the familiar image set in the study phase and were then tested on a surprise recognition memory task.

In all 3 tasks, participants were seated approximately 1m from a cathode ray tube screen. A fixation cross was presented at the center of a black screen for a random time between 800-1200 ms. The stimulus (horizontal or vertical) was then flashed for 20 ms ($7^{\circ} \times 10^{\circ}$ visual angle). Participants had to release their finger as quickly and as accurately as possible if they saw a target stimulus (animal in the categorization task, familiar image in the recognition task) or as soon as they identified the image (study task). If they did not release their finger for 1 s then the response was considered as either a correct rejection if the image was a distractor or a miss if the image was a target. Go response latencies were recorded via a response pad synchronized with the stimulus display.

Training: categorization task

The trained group performed an animal/non-animal categorization on their 200 familiar images set for 15 days. On the two last days, this short training categorization task was followed by a testing session. In this session, participants performed an animal/non-animal categorization task over 12 blocks of 100 trials. The familiar image set was repeated 3 times and mixed with 600 new images. Therefore in total, participants saw their familiar image set 21 times during the training phase. The results of the categorization test showed that 3 weeks training had little effect on categorization performance. Accuracy was marginally increased and the slowest responses were accelerated. However, the fastest reaction times did not change (full details about the procedure and the results of this training phase have been published in Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001).

Study: familiarization with the image set

Before the recognition task, both the trained and the untrained group saw their familiar image set once (which was the first time for the untrained group). The trained group was asked to release their finger as soon as they detected a picture whereas the untrained group was asked to release their finger as soon as they identified the stimulus flashed on the screen. This task had two different purposes. First it familiarized the untrained participants with the set of stimuli to be tested in the recognition task. Second, by asking trained participants to release their finger on all images, we tried to reduce the automatic response behavior (response to animal) that they could have learned during the training phase.

Testing: recognition memory task

Following the study phase, participants of the two groups performed an unexpected recognition task. They were asked to release their finger as soon as they felt that they had seen the same picture earlier. They were told to try to answer on half of the trials. The 200 familiar image set was mixed with 600 new images (different from the ones used in the categorization test). Participants performed 12 blocks of 100 trials, the familiar image set being repeated once every 4 blocks (3 times in total). All images were randomly presented and in each block, 25 images were familiar animal images, 25 were familiar non-animal images, 25 were new animal images and 25 were new non-animal images.

Behavioral analysis

Median reaction times (RT) for correct go responses and d' (Snodgrass & Corwin, 1988) were calculated in the recognition task for each participant and used in statistical analyses. Independent t-tests were applied to compare performance between the trained and the untrained group. Because the familiar image set was repeated 3 times over the experiment (once every 4 blocks), we tested any effect of repetition by dividing the results into 3 (4 blocks x3) and applied a 3 ways repeated measure ANOVA. If an effect was found, post-hoc analysis using paired t-tests between the 1st and 2nd repetition as well as between the 2nd and 3rd repetition were applied. P-values are reported after Bonferroni correction. Finally, as the status of the image during the training phase (animal images were target and non-animal distractors) could influence recognition, we compared performance for animal and non-animal stimuli by using paired t-tests for each of the 2 groups separately. Results are given with the average \pm standard error of the mean (SEM).

EEG recording and pre-processing

For the trained group, EEG activity was recorded using a 32-channel SynAmps amplification system (Neuroscan) using a sampling rate of 1kHz and linked ears as the reference. For the untrained group, a 64-channel BioSemi Active-Two system with Ag/AgCl electrodes (active electrodes) was used, using a sampling rate of 1024Hz. Data were re-referenced off-line to an average reference. Because we did not use the same EEG recording system for both trained and untrained participants, direct comparisons between the 2 groups concerning their EEG activity is impossible (see Figure 3). Therefore each EEG analysis was performed for each group separately.

Data pre-processing was conducted using EEGLAB (Delorme & Makeig, 2004). The EEG signal was band-pass filtered at 0.2-512Hz and a 50Hz notch filter was used. An Independent Component Analysis was applied to identify eye-blinks that were later removed. In a few cases, components isolating obvious muscle artifacts were removed by visual inspection but in general a conservative approach was taken to avoid removing meaningful neuronal activity. The EEG signal was divided into

epochs from -100ms to 600ms relative to stimulus onset. Each epoch was baseline corrected by subtracting the average activity between -100ms to 0ms for each channel.

ERP analysis

For each participant, the average EEG activity for familiar images on one side and new images on the other side was computed (old/new effect). A paired t-test was then applied to compare ERPs in these two conditions for each electrode and at each time point. The obtained p-values were corrected for multiple comparisons using False Discovery Rate (FDR) estimation. Based on these first results and on previous studies showing the important role of frontal and parietal regions in memory tasks (for a review see Rugg & Curran, 2007), we chose two regions of interest: a frontal region including F3, Fz, F4 and a parieto-occipital region including PO3, POz, PO4 (see Figure 3). ERP differences between familiar and new images for these two regions were computed for each time point using FDR-corrected paired t-tests. The same procedure was used to compare the ERPs of animal and non-animal images.

Because an old/new effect was found for the trained group and not for the untrained group, further ERP analyses were only performed for the trained group. The first analysis tested the effect of repetition of the familiar images over the recognition memory task. For the two regions of interest, the old/new effect was computed for each of the repetitions (three times). A repeated measures ANOVA was then applied at each time point with FDR-correction. A second analysis tested the effect of stimulus category on the old/new effect. For each stimulus type (animal or non-animal), an old/new difference was computed (familiar animal – new animal and familiar non-animal – new non-animal) for the two regions of interest. Paired t-tests were then performed with FDR correction.

Studies on recognition memory tasks usually compute ERPs for Hits and correct rejections. However, we decided to use all trials (correct and incorrect responses) and divide the data only depending on the trial condition for several reasons. First, untrained participants performed the recognition task close to chance level (see behavioral results). ERP analyses using only correct trials would have been very noisy since very few trials would be kept and within these trials, some might be correct even though participants did not recognize the image. On the contrary, the trained group performed the task with very high performance so ERP results would not differ whether incorrect trials were included in the analysis or not. Second, including all trials allowed us to test whether the ERP difference between familiar and unfamiliar trials was due to recognition or the repetition of the familiar image set. If a difference is found in both trained and untrained groups then the ERP difference would be interpreted as the consequence of the repetition of the familiar stimuli (priming) and not the neuronal correlates of familiarity. Finally we wanted to be consistent with all analysis performed in this study. In the comparison between animal and non-animal images whether participants recognized the image as familiar or new did not matter thus all trials were included.

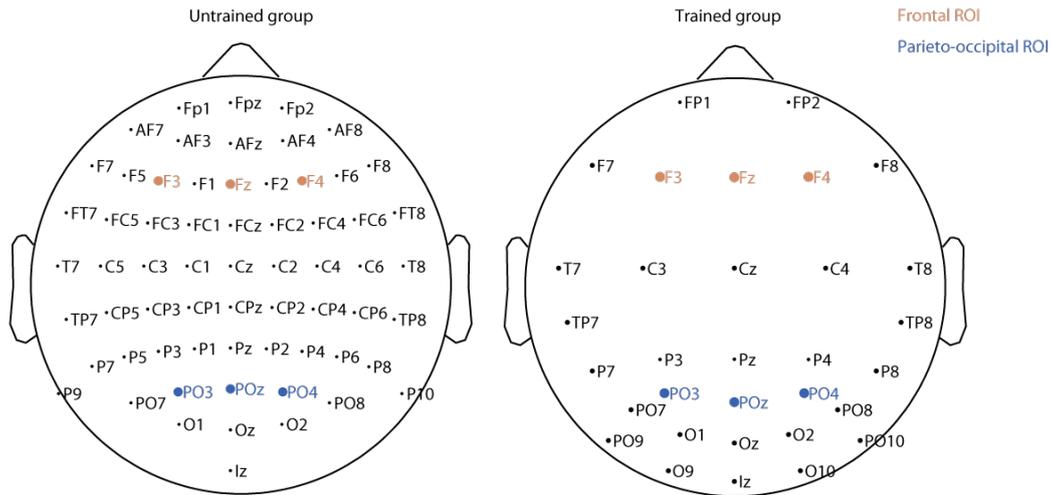


Figure 3. Location of the electrodes for the untrained (left) and trained (right) group. The electrodes included in the two regions of interest are in brown (frontal) and blue (parieto-occipital).

Results

Memory for briefly flashed images

Recognition performance was much higher for trained than for untrained participants ($t(21)=8.52$, $p=3E-8$). Untrained participants could recognize $64.47 \pm 2.46\%$ of the images ($d'=0.83 \pm 0.15$) that they saw in the study phase. However, trained participants were able to perform the memory recognition task at $88.70 \pm 1.49\%$ ($d'=2.67 \pm 0.15$) in 551 ± 8 ms, the best participant recognizing 96.50% ($d'=3.68$) of the images (Figure 4). This suggests that trained participants stored the briefly flashed familiar image set in long term memory while they categorized them in the training phase. Thus, encoding of the images was done without explicit instruction to memorize them. Without such instruction, participants did not look for or try to remember features that would enable them to distinguish between two exemplars of the same category (e.g. specific attributes of a lion that would help recognize it from another lion) but were able to perform the recognition task with a high level of accuracy, suggesting that the implicit representation included many details of the image.

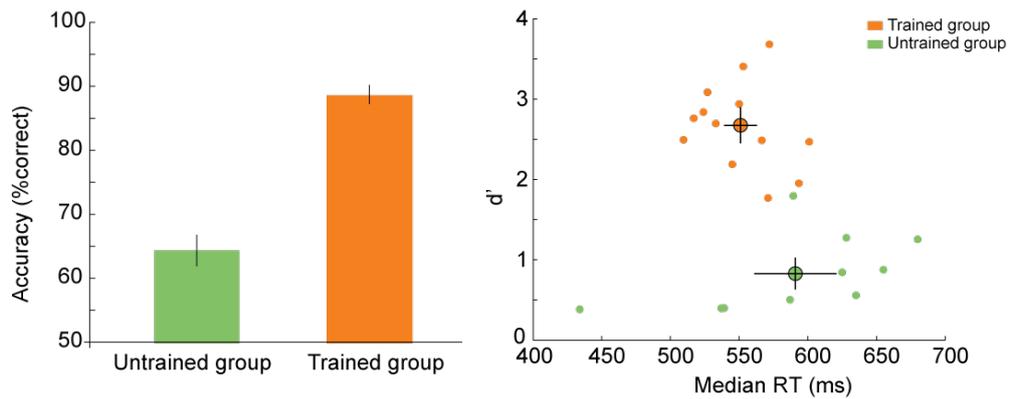


Figure 4. **Behavioral performance in the recognition memory task (familiar/non-familiar).** A. Accuracy (%correct) for untrained (green bar) and trained (orange bar) participants. B. d' as a function of RT. Each dot represents the average performance of one participant. Average performance for each group is represented by a bigger dot with SEM.

To assess the neural correlates of familiarity, we compared participants' ERP for familiar vs. new images for all electrodes at each time point (see method section and Figure 5A). We found no statistical difference between these two conditions for the untrained group. However, a significant difference was found for the trained participants starting around 230ms after stimulus onset. This familiarity effect appeared first in frontal and central regions and later in parieto-occipital and occipital regions. To further investigate the difference between these two regions, we represented ERP for frontal and parieto-occipital regions separately (see method section and Figure 5B and 5C). For the trained group, the familiarity effect was significant from 230ms after stimulus onset in frontal regions and from 315 or 390ms in parieto-occipital regions. This dissociation could be interpreted as the familiarity (early) and recollection (late) components observed in other ERP studies (Curran & Cleary, 2003; Curran, 2000; Vilberg et al., 2006).

One could argue that the ERP difference between familiar and new images reflects a difference in motor response (go response for familiar images vs. no-go response for new images). However, the motor ERP component is usually located around central or parietal electrodes while in our study the effect is seen in all electrodes. Furthermore, if we compare go responses to no-go responses for the untrained participants, we do not find any difference in the first repetition when performance was around chance (see supplementary Figure 1A) and the difference was predominantly after 500ms in parietal regions when considering the entire experiment (see supplementary Figure 1B). Another possibility could be that the familiarity effect was due to the three repetitions of the images in the experiment (performance and ERPs were analyzed separately for each repetition in the next section). However, the effect was present for the trained group but not for the untrained group. Therefore, the ERP difference between familiar and new images found for the trained group was most probably related to memory.

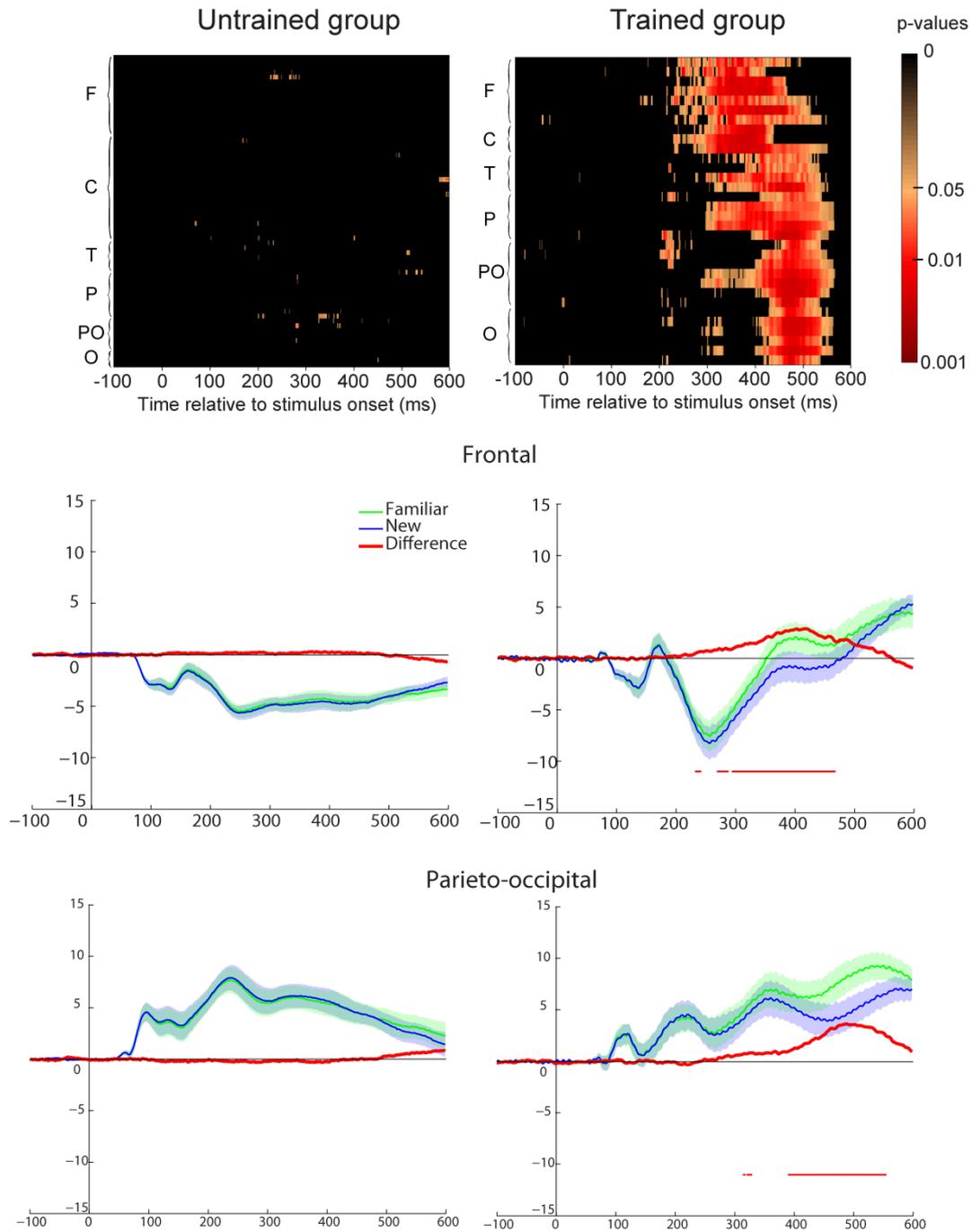
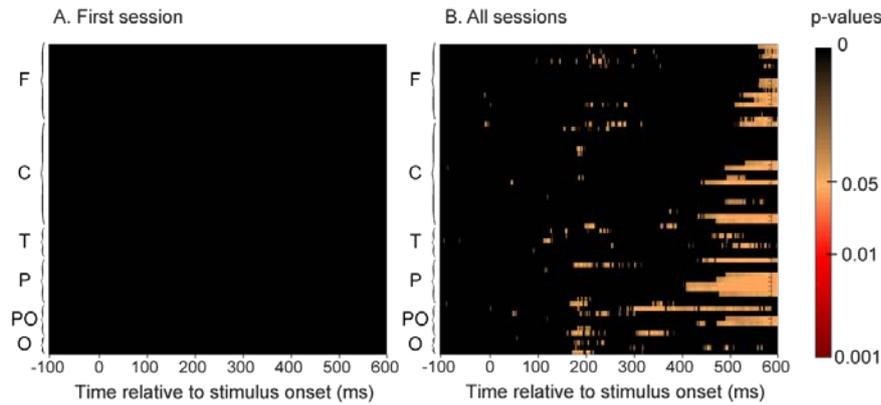


Figure 5. Recognition memory task. **A.** Paired t-test p-values (FDR-corrected) for the ERP difference between all familiar and new images. The difference is represented at each time point from -100 to 600ms relative to stimulus onset. Electrodes are grouped by regions (F: Frontal, C: Central, T: Temporal, P: Parietal, PO: Parieto-occipital, O: Occipital with left hemisphere electrodes represented above right hemisphere electrodes). Untrained group results are on the left and trained group results are on the right. **B.** Frontal ROI. Average ERP for familiar (green) and new (blue) images. The shaded regions represent the standard error of the mean (SEM). The difference between the two ERPs is represented in red. For each time-point a paired t-test was calculated (FDR -corrected) and reported by a dot at $-12\mu\text{V}$ if it was significant ($p < 0.05$). **C.** Parietal ROI. Legend as in B.



Supplementary Figure 1. Paired t-test p-values (FDR-corrected) for the ERP difference between go and no-go responses for the untrained group considering only the first repetition (when accuracy was at chance level) (A) or the entire testing phase (B). The difference is represented at each time point from -100 to 600ms relative to stimulus onset. Electrodes are grouped by regions (see legend Figure 5).

Repetition effect within the experiment

Because the familiar image set was used 3 times during the recognition task, it is possible that some learning occurred over the course of the experiment. In the procedure, the familiar image set was repeated once every 4 blocks. Therefore, to assess the effect of learning we compared participants' performance in each of the 3 repetitions separately for the trained and untrained groups (Figure 6). Considering the untrained group, RTs did not change over the experiment ($F(2,18)=1.18$, $p=0.33$). However we found an effect for d' ($F(2,18)=18.81$, $p=4E-5$, $\eta^2=0.99$). Post-hoc comparisons showed that d' increased between both the first and second repetition ($t(9)=2.92$, $p=0.03$) and the second and third repetition ($t(9)=6.01$, $p=4E-4$). Considering the trained group, we found an effect of the familiar image set repetition for RTs ($F(2,24)=6.98$, $p=0.004$, $\eta^2=0.89$) and d' ($F(2,24)=12.57$, $p=2E-4$, $\eta^2=0.99$). Post-hoc analysis showed that trained participants became faster and more accurate between the first and second repetition ($t(12)=3.15$, $p=0.02$; $t(12)=4.21$, $p=0.002$; for RT and d' respectively) but no difference was found between the second and third repetition ($t(12)=1.24$, $p=0.48$; $t(12)=0.22$, $p=1.66$; for RT and d' respectively). This suggests that trained participants performance might be close to ceiling performance from the second repetition. On average, untrained participants increased their d' recognition performance by 0.67 (from 0.52 ± 0.12 to 1.19 ± 0.21) and trained participants by 0.38 (from 2.47 ± 0.17 to 2.85 ± 0.22) between the first and the last repetition of the familiar images.

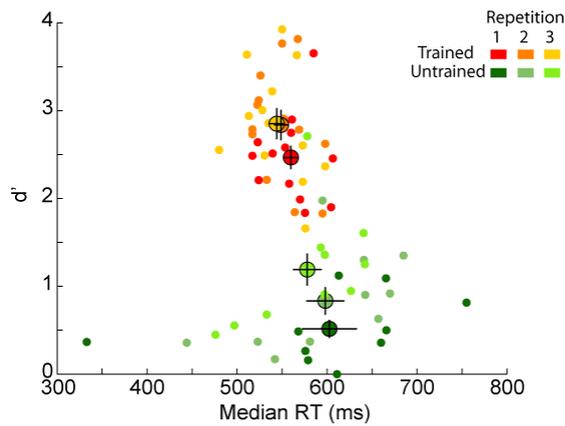


Figure 6. Behavioral results (d' as a function of RT) for each repetition of the familiar image set. Average performance for each group (trained participants in orange, untrained participants in green) is represented by a bigger dot with SEM. Each dot represents the performance of one participant in each of the three repetitions separately (from dark for the first repetition to lighter colors for the subsequent repetitions). Note that although the majority of untrained participants had lower d' performance than trained participants, one untrained participant reached similar d' performance to trained participants in the last repetition.

We found an effect of repetition on d' for both groups of participants. However, whereas learning occurred over the experiment for the untrained group (which could reflect medium term memory, e.g. Melcher, 2001), the same cannot be concluded for the trained group. Indeed, trained participants increased their performance only between the first and second session and for both d' and RTs. Thus this effect might be the consequence of learning a new task (recognition) compared to the task they used to perform in the training stage (categorization). Another hypothesis could be that a memory trace was reactivated after the first repetition of the stimulus. However, because both groups were presented with the stimulus set (study phase) before being tested in the recognition task, this memory trace could have already been reactivated (if needed) at this stage of the experiment.

ERP studies have shown that the early frontal difference between familiar and new items varies linearly with familiarity strength but is not sensitive to different degrees of recollection (Curran & Cleary, 2003; Woodruff, Hayama, & Rugg, 2006; Yu & Rugg, 2010). However, the late parietal component was correlated with the number of correctly retrieved contextual details (Vilberg et al., 2006; Vilberg & Rugg, 2009). Thus, the strength of familiarity is reflected in the early frontal component while the amount of retrieved information is reflected in the late parietal component. To assess any difference in the amount of familiarity or recollection we compared the amplitude difference of the old/new effect between the three repetitions for the frontal and parieto-occipital components that we identified earlier for the trained group. We found no evidence for larger familiarity strength at the end of the experiment (no significant difference between the three repetitions for the frontal ROI). However, even though not significant, we observed a trend for a higher amplitude difference between the first and second repetition in the parieto-occipital ROI that matched behavioural findings. This parieto-occipital effect can be interpreted in three ways. First, since this component is supposed to reflect the

amount of retrieved information, it is possible that participants recalled more details of the images after the first repetition which helped them to perform the task. Alternatively, in the specific context of our experiment, it might reflect a difference in the motor decision or motor command and not recollection processes. Indeed the old/new effect could be the difference between go and no-go responses and this difference would be higher after the first repetition because of the participants' performance improvement. Finally, this trend could be related to both memory and motor components taking place around the same time.

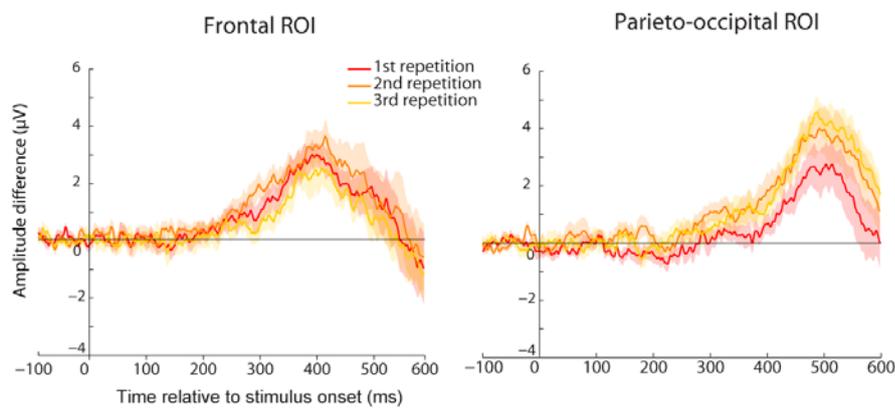


Figure 7. **Average ERP amplitude difference between familiar and new images** for the first (red), second (orange) or third (yellow) repetition of the familiar set. The shaded regions represent the SEM. Results are represented for the trained group either for the frontal (left) or the parieto-occipital (right) ROI. No significant difference was found between the repetitions.

Effect of stimulus status in the training phase

The familiar image set was incidentally learned by the trained participants while they were performing an animal/non-animal task. There are many reasons to think that animal images could have been encoded better than non-animal images. First, as animals were target, participants might have paid more attention to them. Second, target images were linked to a specific concept (animal) while distractor images were not (non-animal images could be houses, city landscape, forests, etc.). For example, Wiseman & Neisser (1974) have shown that participants tested on ambiguous faces recognized stimuli judged as a face during the learning phase better than stimuli judged as a meaningless pattern. Finally, recent studies suggest that memory has evolved such that animate stimuli are better remembered than inanimate ones (Nairne, Pandeirada, & Thompson, 2008; Nairne & Pandeirada, 2010a, 2010b). This “animacy effect” has been found in recognition memory tasks involving words (Nairne, Vanarsdall, Pandeirada, Cogdill, & Lebreton, 2013; VanArsdall, Nairne, Pandeirada, & Blunt, 2013) and pictures (Bonin, Gelin, & Bugaiska, 2013).

To test this effect in our experiment we compared participants' performance for animal and non-animal images (Figure 8A). RTs were similar between animal and non-animal images for the untrained group ($t(9)=0.14$, $p=0.89$) but trained participants responded around 34ms faster to animal images ($t(12)=3.94$, $p=0.002$). Interestingly, whereas RT distribution for non-animal images was delayed by around 30ms for the early responses (<400ms) compared to animal images, at longer latencies the two distributions overlapped (Figure 8B). Recognition performance (d') was better for non-animal than animal images for the trained group ($t(12)=4.94$, $p=3E-4$). One might think that low performance for animal images might be an effect of speed-accuracy trade-off. However, d' was also better for non-animal images for the untrained participants ($t(9)=3.20$, $p=0.01$) even though RTs were the same for both types of images. For the untrained and the trained group, animal images generated more responses in terms of false alarms (FA) ($t(12)=5.40$, $p=2E-4$; $t(9)=4.66$, $p=0.001$; for the trained and the untrained group respectively) and Hits ($t(12)=3.08$, $p=0.01$; $t(9)=2.20$, $p=0.06$; for the trained and the untrained group respectively). Thus, there was an overall bias towards familiar responses for animal images. We develop below possible explanations of these results for a) lower d' on animal images and b) faster RTs on animal images for trained participants compared to non-animal images.

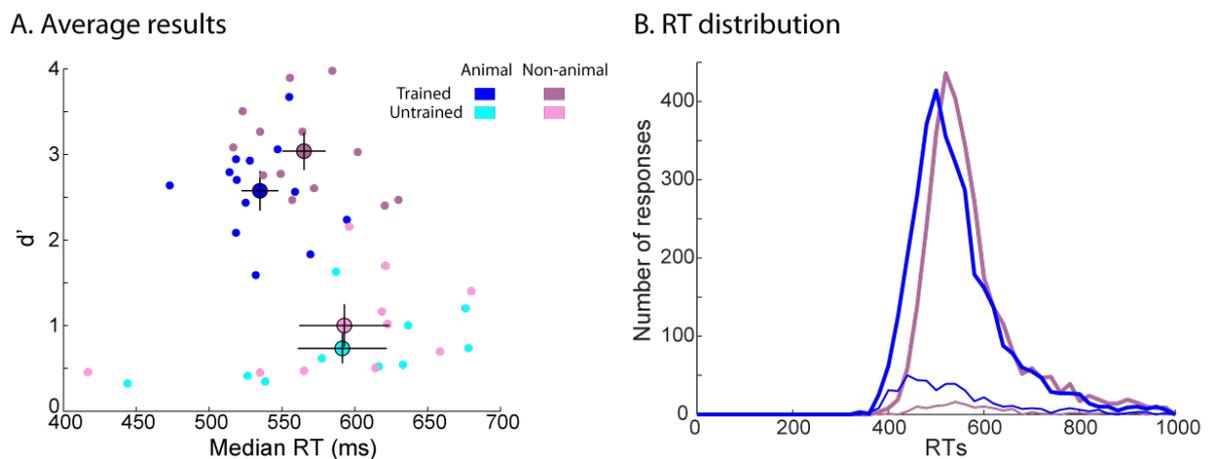


Figure 8. Behavioral results for animal and non-animal images separately in the recognition task. **A.** d' as a function of RT. Each dot represents one participant performance on animal (blue) and non-animal (purple) images. Average performance for each group (trained participants in dark color, untrained participants in light color) is represented by a bigger dot with SEM. **B.** RT distribution for animal (blue) and non-animal (green) images for the trained group. The number of correct (Hit: thick curves) and incorrect (FA: thin curves) responses were computed in a 20ms bin.

Lower d' on animal images

Neither trained nor untrained participants remembered animal images better than non-animal ones. Indeed we found the opposite effect. Compared to studies which found an animacy effect, we tested our participants in a go/no-go task and not in a free recall task. However, we think that the discrepancy might be better explained by the similarity between familiar and new images. It has

previously been shown that the similarity within the familiar image set influences recognition memory performance (e.g. Huebner & Gegenfurtner, 2012; Konkle, Brady, Alvarez, & Oliva, 2010a; Koutstaal & Schacter, 1997). Recognition performance is worse if many stimuli of the same basic category have to be remembered than for very dissimilar stimuli. In our study, the overall number of responses (Hits and FAs) for animal images was higher than for non-animal images so participants were more likely to think that a new image of a given animal (e.g. lion) was familiar since several views of lions were presented during the training phase. Indeed, images within the animal category were more similar than the ones within the non-animal category. In the first case, different kinds of the same basic animal category could be presented whereas in the second case, non-animal images were more diverse (they included plants, fruits, vegetables, fireworks, buildings, vehicles, desert, forest, etc...). In Bonin et al., (2013), the authors ruled out the possibility that differences between animate and inanimate objects was attributable to differences in the richness of perceptual or semantic features. However, they did not control for the amount of similarity between stimuli. Thus, it is possible that in their study animate stimuli were more distinctive than inanimate ones whereas the opposite was true in our experiment.

Faster RTs on animal images for the trained group

Trained participants responded faster to animal than to non-animal images. We computed the ERP old/new effect for the trained group separating animal and non-animal images considering the three repetitions together and then compared these two old/new effects. We performed this analysis on all electrodes and each time point as well as for the two ROIs (Figure 9A). We found a significant difference between the old/new effect for animal and non-animal images mostly visible in parietal, parieto-occipital and occipital regions. The stimulus status (animal or non-animal) had no significant effect on the early old/new effect in frontal regions (Figure 9B) but was significant in parieto-occipital regions. The ERP difference in frontal regions reached higher amplitude for animal images which could be interpreted as higher familiarity strength for animal than for non-animal images. In parieto-occipital regions, the old/new effect started around 30ms later for non-animal images compared to animal images (370ms vs. 405ms). If we consider the parieto-occipital effect as driven by recollection processes, then this shows that recollection of animal images was faster than that for non-animal images. However, we cannot be sure about the nature of this component. Since we used a go/no-go protocol, this effect could also be driven by motor processes. Indeed, this 30ms shift matches that observed between the earliest RT latencies for animal and non-animal images.

Thus, ERP analysis showed that the early stage of visual memory recognition (frontal component) was similar between animal and non-animal images. The old/new effect started with the same latency for both types of image. Later processes in the parieto-occipital region were faster for animal than for non-animal images, possibly because they were targets in the training phase (therefore participants remembered these images better), explaining faster RTs for animal images. These processes could either be explained by a faster access to details about the stimulus in long-term memory (recollection) or faster motor processes. This second explanation could be a consequence of the training procedure in which participants had to release their finger for animal images and inhibit their response for non-

animal images. By repeating this categorization task, a higher fluency for animal images and/or an inhibition for non-animal images might have been set up.

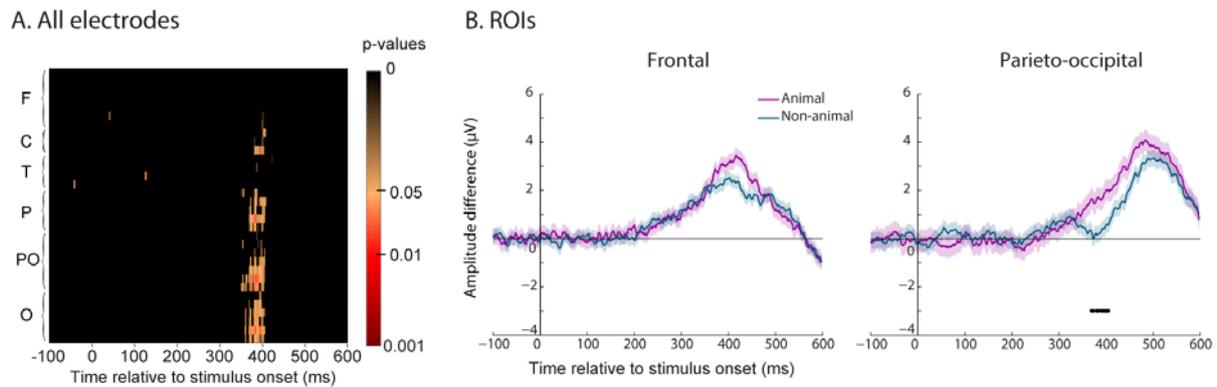


Figure 9. Difference between the ERP old/new effect for animal and non-animal images for the trained group. **A.** P-values (FDR-corrected) of this difference are represented for each electrode grouped by regions (same legend as in Fig.5). **B.** The average amplitude difference of the old/new effect is represented for animal (purple) and non-animal (blue) images for the two ROIs (frontal on the left and parieto-occipital on the right). The shaded regions represent the SEM. Significant difference ($p < 0.05$) between animal and non-animal old/new effect is reported by a dot at $-3\mu\text{V}$.

Automatic processing of object categories

An ERP difference between animal and non-animal images starting around 150ms after stimulus onset has been reported in many animal/non-animal categorization tasks (e.g. Fabre-Thorpe et al., 2001; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). Moreover, Carlson, Tovar, Alink, & Kriegeskorte (2013) have shown that the category of an image can be decoded using the EEG signal even though participants were not performing a categorization task. Here we tested whether we could find an ERP difference between animal and non-animal images while participants performed a familiar/non-familiar task.

For both the trained and untrained participants, we compared the ERP elicited for an animal image to the one elicited for a non-animal image, regardless of whether the image was familiar or new. The results showed a significant difference between these two conditions for both groups of participants (Figure 10A). This category effect was present mainly in frontal and central electrodes and was significant around 170ms after stimulus onset even though it started around 150ms (Figure 10B). Thus the stimulus category, animal or non-animal, was automatically processed while participants performed an unrelated task.

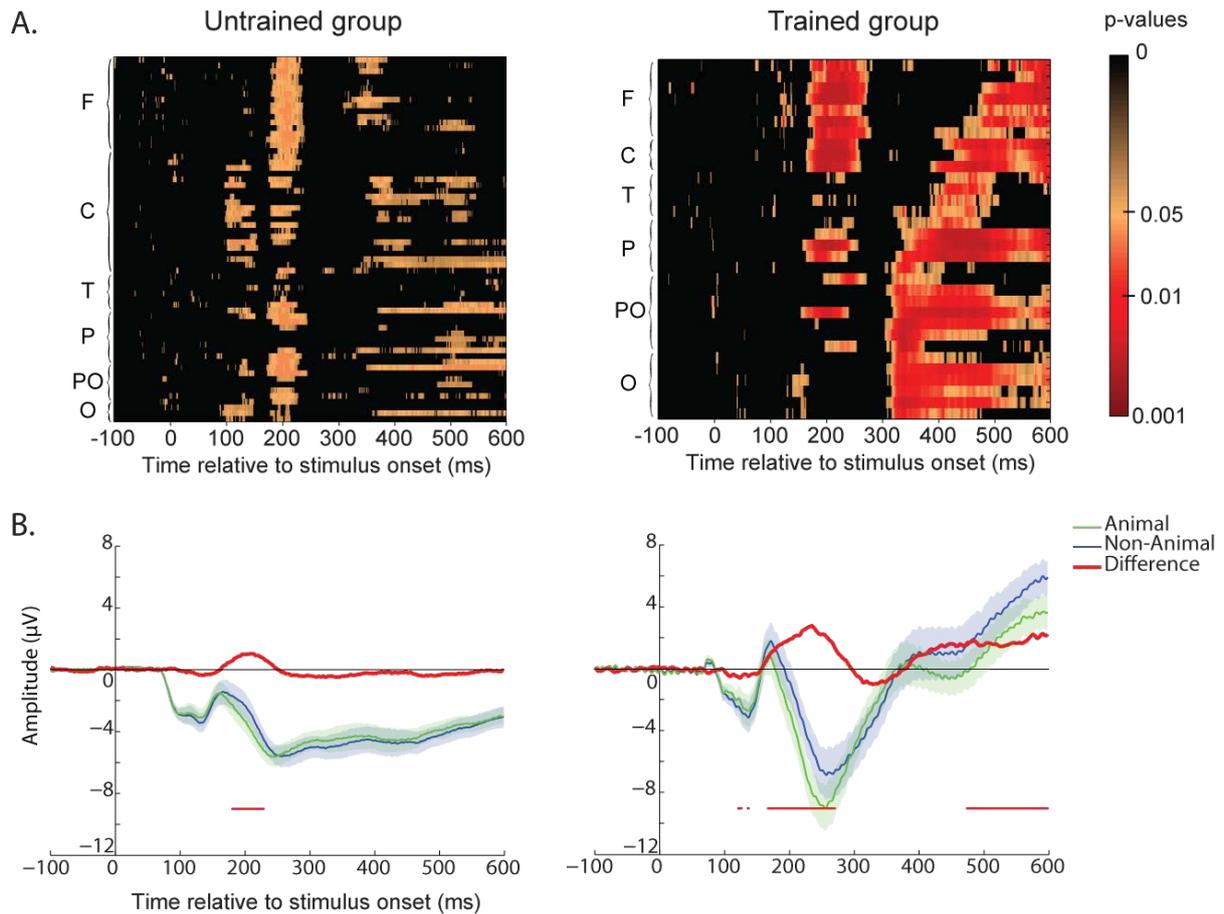


Figure 10. Automatic processing of visual object categories in the recognition memory task. A. Paired t-test p-values (FDR-corrected) for the ERP difference between all animal and non-animal images. Same legend as in Fig.5. B. ERP amplitude for animal and non-animal images for Frontal ROI. The shaded regions represent the SEM. The difference is represented in red with significant p-values ($p < 0.05$) reported by a dot at $-9\mu\text{V}$.

Discussion

The main goal of our study was to test whether 20ms flashed natural scenes could be stored in long-term memory. We tested two groups of participants: one group saw the 200 familiar images set only once and another group saw the same images many times over a period of three weeks. Our results show that after only one viewing of 200 flashed images, participants can only remember 59% of them (in the first repetition) but if the same set of flashed images is repeated over 15 days in an unrelated task, participants can remember 87% of the images. These results are impressive for several reasons. First, trained participants were tested one week after the training phase so any recognized image had to be stored in long-term memory. Second, images were presented for only 20ms, which avoids any possible eye movements during the encoding or the recognition of the images. Third, participants were not explicitly asked to memorize these images and did not know that they would be tested on a recognition task. Fourth, to reach that level of performance, participants had to remember enough

detail of their familiar image set to be able to distinguish between two exemplars of the same basic category (e.g. between an old or a new lion). Fifth, participants were tested in a speeded categorization task (go/no-go paradigm) so they had to retrieve the image information quickly enough to be able to perform the task.

Implicit encoding in long term memory

Trained participants were able to recognize 87% of familiar images. It could be argued that this high level of performance was a consequence of viewing the 200 images a few minutes before the experiment. Indeed, when testing is delayed, recognition performance is usually lower than if the testing happens the same day as the training phase (e.g. Huebner & Gegenfurtner, 2012). However, untrained participants also saw the 200 images a few minutes before the recognition task but their performance were close to chance level (59%). Moreover, both groups of participants were not aware that they would be tested on a recognition task. Thus, even if the presentation of the images before the task induced a reactivation of a memory trace for the trained group, it still means that a memory trace was created during the categorization task performed a week before the recognition task.

In other studies in which pictures were incidentally learnt, recognition performance was much better than what we report here for the untrained participants. For example, using a similar living/non-living categorization task as a study phase, studies reported 70% (Groh-Bordin, Zimmer, & Mecklinger, 2005) or even 87% (Küper, Groh-Bordin, Zimmer, & Ecker, 2012) correctly recognized objects. Even when the new images included new objects from the same category as familiar ones, participants' recognition performance was still very high (e.g. 75% in Ecker & Zimmer, 2009). However, there are two major differences between these kinds of studies and our own. First, even when similar objects were used, only one stimulus per object category was included. Second, images were presented for a much longer time than in our study (from 500ms to 2000ms). Compared to the trained group, these two factors might have allowed better implicit encoding of the stimuli.

Briefly presented images can be memorized

Earlier studies on recognition memory have typically used very long stimulus presentation times for the encoding and the testing phase. The fastest stimulus duration used was probably in the study of Huebner & Gegenfurtner, (2012) in which stimuli were presented for 750ms in the encoding phase with a 1s blank interval, and for 2000ms in the testing phase with no time limit to respond. Here we used images flashed for only 20ms which avoids any possible saccade that could help participants to encode additional details of the stimulus (Loftus, 1972; but see also Melcher & Kowler, 2001). Indeed even when we accumulate the presentation time of the images over the 22 repetitions before testing (440ms), images were still presented for a shorter duration than in Huebner & Gegenfurtner, (2012).

However, the presentation of the images was not followed by a mask. It is thus possible that participants processed the image until the next trial (around 1s later). Indeed, Intraub, (1979) showed that more than stimulus duration, the time between two to-be-remembered pictures played a larger role in recognition memory performance. In her study participants were presented with a stream of 16 pictures (RSVP paradigm) and were then tested on a recognition task. Stimulus onset asynchrony (SOA) varied from 110ms (no blank interval between 2 pictures) to 1500ms while presentation time of the pictures remained 110ms. Participants were able to recognize 92.5% of the pictures with a 1500ms SOA but only 25% of the pictures with a 110ms SOA. In a following experiment, Intraub, (1980) showed that recognition performance was similar for pictures showed for 110ms followed by a 5890ms blank interval (80% correct recognition) and for pictures showed for 6000ms (94% correct recognition).

However, in these experiments, before the presentation of the images, participants knew that they would be tested subsequently on a recognition task. We can imagine that to better memorize them, they would continually encode additional details even after image disappearance, since these were relevant for the later task. In our study, participants had no reason to do so in the animal/non-animal categorization task used in the training phase. Furthermore in the studies of Intraub, participants were tested only a few minutes later (compared to one week in our study) and could rely on medium term memory (Melcher, 2001). Moreover, new images were selected to be very different from the familiar images: *“Distractors that did not bear a close resemblance to any of the stimuli were chosen”* (Intraub, 1980). Thus, the results of these experiments cannot be directly compared to ours. However, our results might fit better in the framework of the total time hypothesis (Bugelski, 1962; Cooper & Pantle, 1967). This theory suggests that the information stored in memory would depend on the total amount of time that a stimulus has been seen in one or several viewings. Our results are important in this framework since they show that this hypothesis could be applicable for stimuli repeatedly flashed over many days.

How detailed are the stored representations?

The similarity of the familiar and the new image sets can influence memory. In recognition memory tasks, many studies have shown that participants were faster and more accurate to reject stimuli that were different to the familiar objects than to reject stimuli that were similar to them (Wiesmann & Ishai, 2008; Yago & Ishai, 2006). For example, increasing the number of scenes from similar semantic domains reduces memory for scene details (Melcher & Murphy, 2011). The effect of conceptual and perceptual similarity on recognition memory has been tested more systematically using pictures of objects (Konkle et al., 2010a) or scenes (Konkle, Brady, Alvarez, & Oliva, 2010b). In these studies, the authors varied the number of exemplars from each stimulus category that had to be remembered. They observed that performance decreased as the number of stored exemplars for each category increased and this effect was not predicted by perceptual distinctiveness of the exemplars but by conceptual

distinctiveness (category) (see also Huebner & Gegenfurtner, 2012 who showed that both factors matter). In all these studies, recognition performance was high. For example, even when 64 scenes from the same category had to be remembered, memory performance was at 76% accuracy (Konkle et al., 2010b). Altogether these results suggest that visual long term memory depends on conceptual and possibly also visual similarity between the different stimuli to be stored.

In our study, we found that participants recognized animal images with lower accuracy performance than non-animal images. As mentioned in the results section, it is very likely that animals were conceptually more similar than non-animal images thus leading to a higher number of false recognitions. Indeed, half of the images were animals and some of the animals were from the same basic category (e.g. different images of lions). Moreover, compared to the high recognition performance observed in the studies mentioned above, untrained participants had very low performance after only a single viewing of the images. These two observations suggest that the image set that we used required participants to be able to recognize enough details of the images to perform the task well. Remembering the gist of the image or the category of the object might not have been sufficient to perform the recognition task with the high accuracy observed for the trained participants.

Automatic object categorization

ERP analyses revealed a difference between animal and non-animal images even though participants were performing a recognition task. We interpreted these results as an automatic access to object categories. This automatic access can be related to studies showing no requirement for attention in processing the category of natural scenes (Li, VanRullen, Koch, & Perona, 2002; Poncet, Reddy, & Fabre-Thorpe, 2012; Rousselet, Thorpe, & Fabre-Thorpe, 2004). Another study using classification methods showed indirectly that information about stimulus category was present in the EEG signal as early as 120-130ms after the onset of the stimulus presentation even when participants' attention was focused on an unrelated task in the centre of the screen while the object was presented in the background (Carlson et al., 2013). In our study, participants were asked to process the image but in order to determine its familiarity not its category. Nevertheless, we still found a significant effect of object category in ERP analyses from 150-170ms after stimulus onset, which is around the same time as reported in other studies (Fabre-Thorpe et al., 2001; VanRullen & Thorpe, 2001). As for the study of Carlson et al., (2013), we cannot be sure that object categories were not consciously processed. However, categorizing the object was irrelevant for the task and if participants were doing so, their reaction times should be longer than in a recognition task in which all stimuli are from the same category. Participants performed the recognition task in around 550ms, which is as fast as in studies using the same go/no-go paradigm with only face stimuli (Barragan-Jason, Lachat, & Barbeau, 2012; Ramon, Caharel, & Rossion, 2011). Altogether, these results argue in favor of an automatic access to object categories.

Neuronal correlates of recognition memory

ERP analyses showed a significant difference between familiar and new stimuli for the trained group. This old/new effect started around 230ms after stimulus onset in frontal regions, and 300-400ms in parieto-occipital regions, consistent with previous studies reporting shorter latencies for pictures than for word recognition (Curran & Cleary, 2003; Schloerscheidt & Rugg, 1997, 2004). Many studies suggest that this differential activation in frontal and parietal cortex reflects two distinct memory processes, one for familiarity and one for recollection (Curran & Doyle, 2011; Curran, 2000; Vilberg & Rugg, 2009; Yonelinas, Otten, Shaw, & Rugg, 2005; Yonelinas, 1994). However, other studies suggest that recollection and familiarity reflect differences in the strength of a common memory trace (Donaldson, Mackenzie, & Underhill, 1996; Dunn, 2004; Gonsalves, Kahn, Curran, Norman, & Wagner, 2005; Squire, Wixted, & Clark, 2007; Wiesmann & Ishai, 2008; Wixted, 2007). Nevertheless, in both hypotheses the prefrontal cortex and the posterior parietal cortex are involved in recognition processes. Indeed, these two regions respond more strongly in fMRI to stimuli judged as old than as new (e.g. Rissman, Greely, & Wagner, 2010).

The meaning of the old/new effect found in ERP studies is still under debate (e.g. Mecklinger, Frings, & Rosburg, 2012; Paller, Lucas, & Voss, 2012). Different authors suggest that this effect might be the consequence of multiple components (e.g. novelty, familiarity, priming) that are difficult to disentangle. For example, Tsivilis, Otten, & Rugg, (2001) suggested that the frontal effect was not a response to stimulus familiarity but to stimulus novelty. On the other hand, other studies argue that the ERP difference could be a consequence of the repetition of the familiar stimuli (conceptual priming) distinct from the neural correlates of familiarity (Paller, Voss, & Boehm, 2007; Voss & Federmeier, 2011; Voss & Paller, 2006; Voss, Schendan, & Paller, 2010). Against this hypothesis, many studies argue that this effect varies with the perceptual overlap between stimuli at study and test (while conceptual processing remains unchanged) and that semantic priming (implicit retrieval) is located in more centro-parietal regions than the familiarity effect (explicit retrieval) (Bridger, Bader, Kriukova, Unger, & Mecklinger, 2012; Ecker, Zimmer, Groh-Bordin, & Mecklinger, 2007; Küper et al., 2012; Lucas, Taylor, Henson, & Paller, 2012).

In our study, ERP were analysed by including all trials. Thus, if there was any priming caused by the repetition of the images we should have been able to see it in both groups of participants. However, we found an old/new effect only for the trained participants. This suggests that the ERP difference between familiar and new images reflects a process present for the trained group and absent in the untrained group. Furthermore, compared to priming studies, reaction times were not faster for familiar than new items in the categorization task performed a week earlier (Fabre-Thorpe et al., 2001). Thus the early frontal component might be taken as an indication of familiarity. Concerning the parieto-occipital component, our results show that it was tightly linked to the speed and accuracy of responses. Because our protocol involved a go response on familiar images and a no-go response on new images, we cannot rule out the possibility that it reflects motor activity rather than recollection processes. Indeed, one could argue that our participants did not recollect the images since they were

only asked to perform a familiar/new task. However, the image set could include very similar objects and such difficult distinctions might therefore not be possible to perform based on only familiarity. It is also possible that both motor and recollection components underlie the parieto-occipital effect that we observed.

Conclusion

Our results shows that detailed information about complex natural scenes repeatedly flashed for only 20ms can be implicitly stored in long term memory. Participants were able to recognize familiar images if they had seen them several times in an unrelated task but not if they had seen them only once. This might be possible because the stimuli were repeated over many days in a categorization task and not just passively viewed. Our study highlights two functions that the brain seems to do without any top-down influence. First, information about object categories seems to be processed even though there is no requirement for it. Second, detailed encoding of visual stimuli happens without any explicit instruction.

References

- Barragan-Jason, G., Lachat, F., & Barbeau, E. J. (2012). How Fast is Famous Face Recognition? *Frontiers in Psychology*, 3, 454. doi:10.3389/fpsyg.2012.00454
- Bonin, P., Gelin, M., & Bugaiska, A. (2013). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Memory & Cognition*. doi:10.3758/s13421-013-0368-8
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325-14329. doi:10.1073/pnas.0803390105
- Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology*, 2(1), 1-3.
- Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage*, 63(3), 1334-1342. doi:10.1016/j.neuroimage.2012.07.047
- Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 63, 409-412.
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, 13(10). doi:10.1167/13.10.1
- Cooper, E. H., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, 68(4), 221-234.
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, 28(6), 923-938.
- Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research*, 15(2), 191-205. doi:10.1016/S0926-6410(02)00192-1
- Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience*, 23(5), 1247-1262. doi:10.1162/jocn.2010.21464
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.
- Donaldson, W., Mackenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review*, 3(4), 486-490. doi:10.3758/BF03214551
- Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological Review*, 111(2), 524-542. doi:10.1037/0033-295X.111.2.524
- Ecker, U. K. H., & Zimmer, H. D. (2009). ERP evidence for flexible adjustment of retrieval orientation and its influence on familiarity. *Journal of Cognitive Neuroscience*, 21(10), 1907-1919.
- Ecker, U. K. H., Zimmer, H. D., Groh-Bordin, C., & Mecklinger, A. (2007). Context effects on familiarity are familiarity effects of context - an electrophysiological study. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 64(2), 146-156. doi:10.1016/j.ijpsycho.2007.01.005
- Endress, A. D., & Potter, M. C. (2012). Early Conceptual and Linguistic Processes Operate in Independent Channels. *Psychological Science*, 23(3), 235-245. doi:10.1177/0956797611421485

- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience*, 13(2), 171–180.
- Gonsalves, B. D., Kahn, I., Curran, T., Norman, K. A., & Wagner, A. D. (2005). Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron*, 47(5), 751–761. doi:10.1016/j.neuron.2005.07.013
- Groh-Bordin, C., Zimmer, H. D., & Mecklinger, A. (2005). Feature binding in perceptual priming and in episodic object recognition: evidence from event-related brain potentials. *Cognitive Brain Research*, 24(3), 556–567. doi:10.1016/j.cogbrainres.2005.03.006
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136. doi:10.1037//0096-1523.28.1.113
- Huebner, G. M., & Gegenfurtner, K. R. (2012). Conceptual and Visual Features Contribute to Visual Memory for Natural Images. *PLoS ONE*, 7(6), e37575. doi:10.1371/journal.pone.0037575
- Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2), 78–87. doi:10.1037/0278-7393.5.2.78
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 6(1), 1–12.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558–578. doi:10.1037/a0019165
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychological Science*, 21(11), 1551–1556. doi:10.1177/0956797610385359
- Koutstaal, W., & Schacter, D. L. (1997). Gist-Based False Recognition of Pictures in Older and Younger Adults. *Journal of Memory and Language*, 37(4), 555–583. doi:10.1006/jmla.1997.2529
- Küper, K., Groh-Bordin, C., Zimmer, H. D., & Ecker, U. K. H. (2012). Electrophysiological correlates of exemplar-specific processes in implicit and explicit memory. *Cognitive, Affective & Behavioral Neuroscience*, 12(1), 52–64. doi:10.3758/s13415-011-0065-7
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596–9601.
- Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 3(4), 525–551.
- Lucas, H. D., Taylor, J. R., Henson, R. N., & Paller, K. A. (2012). Many roads lead to recognition: Electrophysiological correlates of familiarity derived from short-term masked repetition priming. *Neuropsychologia*, 50(13), 3041–3052. doi:10.1016/j.neuropsychologia.2012.07.036
- Martini, P., & Maljkovic, V. (2009). Short-term memory for pictures seen once or twice. *Vision Research*, 49(13), 1657–1667. doi:10.1016/j.visres.2009.04.007
- Mecklinger, A., Frings, C., & Rosburg, T. (2012). Response to Paller et al.: the role of familiarity in making inferences about unknown quantities. *Trends in Cognitive Sciences*, 16(6), 315–316. doi:10.1016/j.tics.2012.04.009
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, 412(6845), 401–401. doi:10.1038/35086646

- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6(1).
- Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, 41(25–26), 3597–3611. doi:10.1016/S0042-6989(01)00203-6
- Melcher, D., & Murphy, B. (2011). The role of semantic interference in limiting memory for the details of visual scenes. *Frontiers in Psychology*, 2, 262. doi:10.3389/fpsyg.2011.00262
- Nairne, J. S., & Pandeirada, J. N. S. (2010a). Adaptive memory: ancestral priorities and the mnemonic value of survival processing. *Cognitive Psychology*, 61(1), 1–22. doi:10.1016/j.cogpsych.2010.01.005
- Nairne, J. S., & Pandeirada, J. N. S. (2010b). Adaptive memory: nature’s criterion and the functionalist agenda. *The American Journal of Psychology*, 123(4), 381–390.
- Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: the comparative value of survival processing. *Psychological Science*, 19(2), 176–180. doi:10.1111/j.1467-9280.2008.02064.x
- Nairne, J. S., Vanarsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & Lebreton, J. M. (2013). Adaptive memory: the mnemonic value of animacy. *Psychological Science*, 24(10), 2099–2105. doi:10.1177/0956797613480803
- Paller, K. A., Lucas, H. D., & Voss, J. L. (2012). Assuming too much from ‘familiar’ brain potentials. *Trends in cognitive sciences*, 16(6), 313–315.
- Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Sciences*, 11(6), 243–250. doi:10.1016/j.tics.2007.04.002
- Poncet, M., Reddy, L., & Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision*, 12(1), 15. doi:10.1167/12.1.15
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, 2(5), 509–522.
- Potter, M. C. (2012). Recognition and Memory for Briefly Presented Scenes. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00032
- Potter, M. C., Staub, A., Rado, J., & O’Connor, D. H. (2002). Recognition memory for briefly presented pictures: the time course of rapid forgetting. *Journal of Experimental Psychology. Human Perception and Performance*, 28(5), 1163–1175.
- Ramon, M., Caharel, S., & Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception*, 40(4), 437–449. doi:10.1068/p6794
- Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences*, 107(21), 9849–9854. doi:10.1073/pnas.1001028107
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Research*, 44(9), 877–894. doi:10.1016/j.visres.2003.11.014
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251–257. doi:10.1016/j.tics.2007.04.004
- Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, 392(6676), 595–598. doi:10.1038/33396

- Schloerscheidt, A. M., & Rugg, M. D. (1997). Recognition memory for words and pictures: An event-related potential study. *NeuroReport*, 8(15), 3281–3284.
- Schloerscheidt, A. M., & Rugg, M. D. (2004). The impact of change in stimulus format on the electrophysiological indices of recognition. *Neuropsychologia*, 42(4), 451–466. doi:10.1016/j.neuropsychologia.2003.08.010
- Shaffer, W., & Shiffrin, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology*, 92(2), 292–296. doi:10.1037/h0032076
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 156–163. doi:10.1016/S0022-5371(67)80067-7
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology. General*, 117(1), 34–50.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews. Neuroscience*, 8(11), 872–883. doi:10.1038/nrn2154
- Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2), 207–222. doi:10.1080/14640747308400340
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73–74.
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, 32(5), 579–592.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.
- Tsivilis, D., Otten, L. J., & Rugg, M. D. (2001). Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron*, 31(3), 497–505.
- VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: animacy processing produces mnemonic advantages. *Experimental Psychology*, 60(3), 172–178. doi:10.1027/1618-3169/a000186
- VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4), 454–461.
- Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, 1122(1), 161–170. doi:10.1016/j.brainres.2006.09.023
- Vilberg, K. L., & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping*, 30(5), 1490–1501. doi:10.1002/hbm.20618
- Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, 54(4), 298–303.
- Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, 48(4), 532–546. doi:10.1111/j.1469-8986.2010.01085.x
- Voss, J. L., & Paller, K. A. (2006). Fluent Conceptual Processing and Explicit Memory for Faces Are Electrophysiologically Distinct. *The Journal of Neuroscience*, 26(3), 926–933. doi:10.1523/JNEUROSCI.3931-05.2006
- Voss, J. L., Schendan, H. E., & Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *NeuroImage*, 49(3), 2879–2889. doi:10.1016/j.neuroimage.2009.09.012

- Wiesmann, M., & Ishai, A. (2008). Recollection- and familiarity-based decisions reflect memory strength. *Frontiers in Systems Neuroscience*, 2, 1. doi:10.3389/neuro.06.001.2008
- Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *The American Journal of Psychology*, 87(4), 675-681. doi:10.2307/1421974
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152-176. doi:10.1037/0033-295X.114.1.152
- Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, 1100(1), 125-135. doi:10.1016/j.brainres.2006.05.019
- Yago, E., & Ishai, A. (2006). Recognition memory is modulated by visual similarity. *NeuroImage*, 31(2), 807-817. doi:10.1016/j.neuroimage.2005.12.014
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1341-1354.
- Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(11), 3002-3008. doi:10.1523/JNEUROSCI.5295-04.2005
- Yu, S. S., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research*, 1320, 74-84. doi:10.1016/j.brainres.2009.12.071
- Zimmer, H. D., & Ecker, U. K. H. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neuroscience & Biobehavioral Reviews*, 34(7), 1066-1079. doi:10.1016/j.neubiorev.2010.01.014

4.3. Discussion

Nos résultats montrent que des informations détaillées sur des images flashées plusieurs fois pendant seulement 20 ms dans une tâche de catégorisation visuelle rapide pouvaient être encodées en mémoire à long terme. Malgré ce temps de présentation très court, les stimuli ont été traités jusqu'à un niveau approfondi puisqu'une représentation en mémoire a été créée. Par ailleurs, comme mentionné et ensuite discuté dans l'article 3, ces résultats mettent en évidence un encodage en mémoire à long terme de manière implicite ainsi qu'un rappel assez rapide d'un sentiment de familiarité (ou d'informations sur ces images) même si les réponses des participants étaient beaucoup plus lentes que pour des tâches de catégorisation (environ 200 ms de plus).

L'article dans sa forme actuelle est en préparation et demande certaines révisions. En particulier, il serait plus juste d'utiliser une méthode de groupe (« cluster ») pour corriger pour les comparaisons multiples en ERPs, au lieu d'une méthode basée sur le taux de fausse reconnaissance. De plus, les résultats des tests statistiques obtenus pour les comparaisons des ERPs pourraient être représentés avec la taille de l'effet ou les valeurs de t à la place des valeurs de p . Une analyse intéressante pourrait aussi être ajoutée à cette étude. Pendant l'entraînement sur la tâche animal/non-animal, les performances de catégorisation semblent avoir été améliorées pour les images difficiles. Les participants pourraient donc avoir appris ces images et ainsi améliorer leurs performances pour ces images. Par contre, ils auraient beaucoup moins bien mémorisé les images plus faciles. La comparaison des performances de reconnaissance des images pour lesquelles les participants réalisaient le plus d'erreurs dans la tâche animal/non-animal avec celles des images catégorisées très rapidement et sans erreur dans la même tâche, on pourrait déterminer si les performances de reconnaissance sont liées aux informations pertinentes pour la réalisation de tâche animal/non-animal dans laquelle la mémorisation a eu lieu.

En plus des points de discussion abordés dans l'article, j'aimerais faire quelques remarques sur les limitations du protocole ainsi que quelques suggestions en vue d'une future étude. A la vue des résultats de la première étude visant à améliorer les performances des participants en catégorisation visuelle rapide, Fabre-Thorpe et al. (2001) se sont demandé si les participants avaient mémorisé les images pendant l'entraînement ou pas. Le protocole qui a été utilisé a été construit « après coup » et s'intègre difficilement dans une thématique de recherche. Par exemple, si on s'intéresse au temps de

présentation minimum pour reconnaître des stimuli flashés (e.g. les travaux en RSVP de Potter, 2012) alors il aurait fallu utiliser des stimuli masqués. Si maintenant on s'intéresse seulement à la capacité de mémorisation de stimuli flashés (sans masque), c'est-à-dire sans mouvements oculaires, alors Intraub (1979, 1980) a montré qu'il fallait au moins une seconde pour qu'un stimulus soit mémorisé. Une dernière possibilité serait d'intégrer ces travaux dans les études testant les capacités de mémoire à long terme pour des scènes naturelles mais là encore, le nombre d'images utilisé (200) est beaucoup plus faible que dans ce type d'études (2500 images ont été mémorisées dans l'expérience de (Brady, Konkle, Alvarez, & Oliva, 2008). De plus, au contraire d'études comme celle de Konkle, Brady, Alvarez, & Oliva (2010), la similarité entre les différentes images n'a pas été quantifiée et le niveau de détail des représentations en mémoire n'est donc pas clair.

D'autres paramètres de l'expérience auraient pu être mieux contrôlés et ainsi répondre à certaines questions discutées dans l'article :

- du fait de la procédure go/no-go, il est difficile de distinguer les réponses motrices des corrélats neuronaux de la familiarité. Est-ce que la différence entre les ERPs évoqués pour les images familières et ceux évoqués pour les images nouvelles reflète les corrélats neuronaux du sentiment de familiarité ou sont une conséquence de la préparation motrice ? Ce problème aurait pu être évité en utilisant un paradigme de yes/no, dans lequel une réponse motrice est réalisée quelle que soit l'image présentée. Les participants ont en général des réponses plus lentes que pour des réponses en go/no-go ce qui ne permet pas de connaître les temps de réponses les plus rapides pour réaliser la tâche de familiarité. Mais est-ce vraiment utile de connaître les temps de réaction les plus rapides ? Par ailleurs des réponses plus lentes permettraient aussi de mieux différencier les activités neuronales liées à une réponse motrice de celles liées au traitement du stimulus.
- Le set d'images familières est présenté aux participants entraînés juste avant qu'ils soient testés. La raison donnée est que puisqu'on leur demandait de répondre à toutes les images, cela permettrait d'éviter l'automatisme du « go » sur les images animal et « no-go » sur les images non-animal. Dans ce cas, pourquoi utiliser le set d'images familières et pas un nouveau set d'images ? Ainsi, il aurait été possible de tester les performances de reconnaissance sans pré-activation de la trace mnésique qui a pu être créée. Par ailleurs, on ne peut pas complètement

écarter la possibilité que lorsque les participants revoyaient le set d'images familières, ils aient pu l'encoder de manière beaucoup plus explicite (une tâche de détection étant très simple ils auraient très bien pu porter leur attention sur des détails des images, chose qu'ils n'auraient peut-être pas faite dans une tâche de catégorisation).

- Les 200 images familières étaient répétées trois fois au cours de l'expérience alors que les images nouvelles ne l'étaient jamais. On observe une amélioration des performances de reconnaissance pour les participants non-entraînés au cours de ces trois répétitions. Ce résultat montre qu'il est possible de reconnaître explicitement quelles images sont répétées et lesquelles ne le sont pas. Afin d'éviter un possible rôle de l'apprentissage avec l'augmentation du nombre de répétitions, il aurait été beaucoup plus judicieux de répéter les images familières mais aussi les images nouvelles (n'avoir que deux sets de 200 images répétés trois fois).
- D'après nos résultats, des images présentées pendant seulement 20 ms peuvent être mémorisées (c'est le cas des participants entraînés mais aussi pour la majorité des participants non-entraînés). D'après l'hypothèse du temps total (Bugelski, 1962; Cooper & Pantle, 1967), le nombre d'information encodé dépend du temps de présentation quelle que soit la façon dont ce temps est distribué. Appliquée à notre étude, cela revient à dire que pour mémoriser les images familières, les participants ont besoin d'au moins trois répétitions, c'est à dire 60 ms dans le cas d'un participant non-entraîné, à 22 répétitions c'est à dire 440 ms pour les participants entraînés. Cependant, étant donné que les images ont été présentées sans masque, une stricte multiplication *temps de présentation x nombre de répétitions* ne peut pas s'appliquer. A chaque présentation, les participants pouvaient garder le stimulus en mémoire de travail jusqu'à la présentation de la prochaine image afin d'encoder plus de détails. Ce pourrait être le cas en particulier lors de la tâche de familiarité mais aussi lors de la présentation des 200 images avant la tâche de familiarité (étant donné que les participants devaient répondre à chaque image, tâche très facile et peut être ennuyeuse, ils pourraient plus facilement avoir tendance à visualiser l'image après avoir répondu).

Depuis l'année où les participants ont été entraînés et testés dans la tâche de reconnaissance, plusieurs articles ont été publiés ce qui explique en partie les quelques écueils du protocole et aussi pourquoi il ne s'inscrit pas dans une problématique spécifique. Néanmoins, cette étude est intéressante aussi pour

cette raison puisqu'elle mêle mémorisation implicite à long terme, temps de présentation brefs, rôle de la répétition, et temps de réponse rapides. D'autres études seront nécessaires pour comprendre plus en détail le rôle de ces différents facteurs dans la mémoire à long terme.



Chapitre 5
Interactions



5. Interactions entre les catégories visuelles

La question de l'interaction entre les catégories visuelles développée dans ce chapitre est née de ma lecture de l'article de Evans et al. (2011) qui m'a beaucoup intriguée. Dans leur article les auteurs ont observé des interactions entre les catégories qui étaient dépendantes de la tâche. Dans une de leurs expériences, ils demandaient à des participants d'indiquer si une catégorie cible était présente dans la scène visuelle présentée au sein d'un flot de plusieurs masques. La catégorie cible pouvait être indiquée aux participants avant ou après l'essai. Dans le premier bloc, seulement trois catégories étaient utilisées comme cible dans les différents essais alors que dans le deuxième bloc, six catégories étaient utilisées. Les auteurs ont alors comparé les performances des participants pour des images contenant une catégorie cible (e.g. plage) et une catégorie non-cible (e.g. animal) dans le bloc 1 avec celles où les deux catégories étaient utilisées comme cibles dans des essais différents dans le bloc 2 (e.g. dans le premier essai du bloc 2 les participants indiquaient si l'image contenait un animal et dans le second essai si l'image contenait une plage). Pour ces images contenant deux catégories, les performances diminuaient nettement entre le bloc 1 (lorsqu'une seule des deux catégories pouvait être une cible) et le bloc 2 (lorsque les deux catégories pouvaient être utilisées comme cible dans des essais différents). Deux catégories cibles dans un même bloc d'essais mais pas dans le même essai interfèrent donc entre elles si elles sont présentes dans une même image.

Ces résultats m'ont donné l'idée que les performances pourraient être influencées suivant le contexte de l'essai (quelle était la catégorie cible dans l'essai précédent) et ainsi expliquer les différences entre des tâches de catégorisation utilisant la même catégorie cible pendant plusieurs essais et les tâches de vérification dans lesquelles la catégorie cible change à chaque essai (voir chapitre 2.3.3). De plus, il m'a semblé important de tester non pas l'influence du contexte de la tâche (si une deuxième catégorie est aussi parfois considérée comme cible ou non), mais l'influence en général d'une catégorie d'objet sur le traitement d'une autre catégorie présentée par la suite. En effet, l'influence de la tâche sur le traitement des stimuli est probablement liée à l'activité du cortex préfrontal. Cependant, les catégories d'objets sont déjà structurées de manière hiérarchique dans IT (voir chapitre 1.1.3). Par conséquent, le traitement de catégories plus ou moins distantes (par exemple un oiseau et un chien sont des catégories plus proches qu'un oiseau et une voiture) pourrait avoir des effets différents sur la catégorisation d'un second objet visuel. L'utilisation de différentes catégories d'objets permettrait ainsi

de tester comment la similarité des patterns d'activité cérébrale entre deux objets visuels influence leur interaction.

5.1. Présentation de la problématique

L'objectif de l'article 4 est de tester l'interaction entre des catégories d'objets visuels plus ou moins distantes mais simultanément actives. A mon grand étonnement, je n'ai pas trouvé d'article ayant testé l'influence d'une catégorie visuelle sur une autre pour de courts intervalles de temps. Pourtant, il semble tout à fait plausible que le traitement d'un objet visuel puisse être influencé par un autre stimulus si les deux objets sont présentés au sein d'une fenêtre temporelle courte. En effet, j'ai montré précédemment que l'accès à la catégorie d'un objet est rapide (article 1), automatique (article 3) et ne nécessite pas de ressources attentionnelles (article 2). Par ailleurs, les neurones de IT, au sein desquels l'information de la catégorie des objets est présente, continuent de répondre à un stimulus visuel longtemps après sa présentation (Rolls & Tovée, 1994).

5.1.1. Etudier les interactions entre des représentations...

Le protocole qui m'a semblé le plus à même de répondre à cette question est un protocole d'amorçage. Ce terme « amorçage » est utilisé dans différents sens : il peut faire référence à un amorçage réalisé à partir de la lecture d'une liste de mots avant une expérience, un amorçage dû à la répétition d'un même stimulus au sein d'une expérience, etc. Par exemple, des expériences en IRMf ont testé la diminution de l'activité de certaines aires cérébrales (effet d'adaptation) avec la répétition de la présentation d'objets appartenant à la même catégorie (Fairhall, Anzellotti, Pajtas, & Caramazza, 2011; Koutstaal et al., 2001; Pourtois, Schwartz, Spiridon, Martuzzi, & Vuilleumier, 2009; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003; Weiner, Sayres, Vinberg, & Grill-Spector, 2010). Le protocole d'amorçage auquel je fais ici allusion est la présentation d'une image distracteur (amorce) juste avant la présentation d'une image cible pour laquelle les participants doivent réaliser une tâche. Puisque je m'intéresse à l'interaction entre des catégories d'objets simultanément actives, le temps entre l'amorce et la cible doit rester court.

Ce type de paradigme a été utilisé mais avec des amorces qui n'étaient pas perçues de façon consciente (e.g. Dehaene et al., 1998) : l'amorce était présentée très brièvement et apparaissait au milieu de deux masques. Les résultats de ces études ont mis en évidence un effet de congruence : les réponses des

participants étaient plus rapides et/ou plus précises quand l'amorce et la cible induisaient la même réponse que lorsqu'elles induisaient deux réponses différentes. Par exemple, dans une tâche de catégorisation de nombres pairs vs. impairs, les performances étaient meilleures lorsque l'amorce et la cible étaient toutes deux des nombres pairs, par rapport à la condition où l'amorce était un nombre impair et la cible un nombre pair.

L'interaction entre deux stimuli a été beaucoup plus étudiée dans le domaine du langage, en particulier en utilisant des paradigmes de « picture word interference » (Lupker, 1979; Rosinski, 1977). Une image et un mot sont présentés au même moment et le participant doit nommer l'image et ignorer le mot (Figure 25). Ces travaux ont pu montrer un effet de congruence comme celui trouvé dans les paradigmes d'amorçages masqués mais aussi un effet de la relation entre l'image et le mot. Par exemple dans l'article de Costa, Alario, & Caramazza, (2005) si le mot était une partie de l'image (e.g. l'image d'un cerveau avec le mot neurone), alors le temps pour nommer l'image était plus rapide que si l'image et le mot n'avaient rien de commun (e.g. l'image d'un cerveau avec le mot temple). Un autre point important de cette étude est qu'au contraire, si l'image et le mot appartenaient à la même catégorie (e.g. l'image d'un cerveau et le mot estomac), alors les temps de dénomination étaient plus lents que si le mot appartenait à une catégorie différente de l'image (e.g. l'image d'un cerveau et le mot temple).

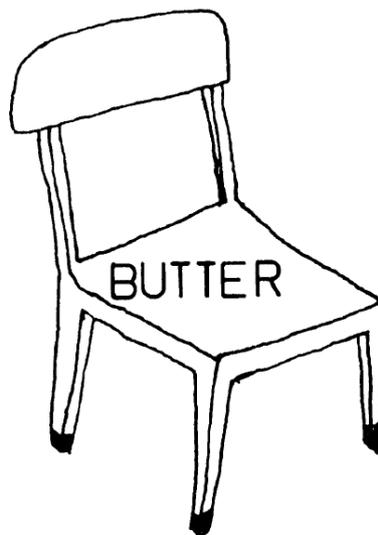


Figure 25. Exemple de stimulus utilisé dans un paradigme d'interférence image-mot. (Lupker, 1979)

Une explication de ces résultats ainsi que ceux d'autres expériences a ensuite été proposée par Mahon, Costa, Peterson, Vargas, & Caramazza, (2007). Selon les auteurs, l'effet du mot distracteur dépend en premier de la tâche : si le mot et l'image convergent vers la même réponse, alors les performances sont améliorées (comme c'est le cas par exemple de l'image d'un cerveau avec le mot neurone). Dans le cas contraire (par exemple l'image d'un cerveau avec le mot estomac), les performances sont détériorées. Si les mots utilisés ont le même degré de convergence avec l'image, alors plus le mot est proche sémantiquement de l'image, plus les participants sont rapides pour nommer l'objet. Par exemple, les participants sont plus rapides pour nommer l'image d'un cheval lorsqu'il est présenté simultanément avec le mot zèbre qu'avec le mot baleine. Pourtant, avec les mêmes stimuli, les participants sont plus rapides pour déterminer que l'image et le mot sont différents si le mot est sémantiquement éloigné de l'image. Par exemple, ils sont plus rapides pour déterminer que cheval et baleine sont différents que pour déterminer que cheval et zèbre sont différents (Mahon et al., 2007).

Ces résultats sont intéressants à prendre en compte puisque l'article 4 a pour objectif de tester l'interaction entre différentes catégories d'objets qui sont plus ou moins proches entre elles. Cependant, nous n'adressons pas cette question dans le domaine du langage (production verbale) mais dans celui de la vision et plus particulièrement en rapport avec l'organisation des catégories visuelles dans IT.

5.1.2. ...en relation avec l'organisation des catégories dans IT

Différents résultats suggèrent que les catégories d'objets seraient représentées dans un espace continu dans IT (voir chapitre 1.1.3). En effet, le degré de similarité entre des patterns d'activité évoqués par différentes catégories d'objets correspond aux similarités intuitives que nous utilisons (Carlson, Simmons, Kriegeskorte, & Slevc, 2014; Kiani et al., 2007; Kriegeskorte et al., 2008; Mur et al., 2013). Par exemple, les patterns d'activité pour les catégories chat et oiseau sont plus similaires que ceux évoqués pour les catégories voiture et oiseau. Ainsi, IT serait organisé de façon hiérarchique et le pattern d'activité évoqué par un stimulus serait plus ou moins similaire à celui d'un autre objet suivant la similarité existante entre ces deux stimuli.

L'objectif de l'article 4 est de tester l'interaction entre des catégories proches (chat et oiseau) et des catégories plus éloignées (voiture et oiseau) lorsque toutes deux sont actives dans IT. Au niveau des populations neuronales, cela revient à tester l'interaction entre deux patterns d'activité similaires et deux patterns d'activité plus éloignés.

5.1.3. Présentation de l'article 4

L'article 4 utilise un protocole expérimental simple pour tester comment les catégories visuelles interagissent quand elles sont actives simultanément. Cette approche n'a jamais été utilisée auparavant mais semble parfaitement adaptée pour étudier l'influence de l'organisation des catégories au niveau cérébral dans la reconnaissance visuelle.

Plus spécifiquement, nous avons utilisé un paradigme d'amorçage, une première scène naturelle (amorçage) était présentée juste avant la présentation de l'image cible pour laquelle les participants devaient réaliser une tâche de catégorisation. Dans l'expérience 1 les participants devaient déterminer si la cible était un oiseau ou non et dans l'expérience 2 si la cible était un animal ou non (réponse oui/non). Afin de limiter l'influence du langage et ainsi tester l'interaction entre les catégories visuelles au début de leur traitement dans IT, les réponses étaient données manuellement (contrairement aux paradigmes d'interférence image-mot, il n'y avait pas de production verbale et aucun mot n'était présenté à l'écran) et les deux images étaient présentées dans un intervalle de temps très court.

Le point important de cette étude est que dans les deux expériences, l'amorce pouvait être soit (a) de la même catégorie que l'image cible, (b) sémantiquement proche de l'image cible, (c) sémantiquement éloignée de l'image cible. Par exemple, dans la tâche de catégorisation oiseau/non-oiseau, l'amorce précédent l'image d'un oiseau pouvait être soit (a) un autre oiseau, (b) un autre animal, (c) un véhicule. En utilisant ces trois types d'amorces, il nous est alors possible de tester l'influence du lien sémantique (ou en termes neuronaux, de la similarité des patterns d'activité) entre l'amorce et la cible sur les performances de catégorisation des participants.

5.2. Article 4: The clash of visual categories³

³ Authors: Ramakrishna Chakravarthi, Michele Fabre-Thorpe, Marlene Poncet (alphabetic order). MP designed the study, ran participants and analysed the data; MP and RC interpreted the data; MP wrote the first draft; RC and MFT gave comments on the manuscript.

The clash of visual categories

Abstract

Categorization is very efficient. However it is unclear how multiple simultaneously active visual categories interact. Such interactions were tested in categorization tasks in which participants ignored a flashed prime, but responded to a 100 ms target presented after an SOA of either 80 or 180ms. The prime's category was either the same as the target category (congruent condition) or different (incongruent condition). In both experiments 1 (bird/non-bird categorization) and 2 (animal/non-animal categorization), responses to congruent trials were faster and more accurate than those to incongruent trials. More importantly, we tested two different types of incongruent primes: *related* and *unrelated* to the target. In Exp1, related incongruent primes belonged to the same superordinate category (e.g. dog), and unrelated ones did not (e.g. car). Interestingly, related incongruent primes caused more interference than unrelated ones at both SOAs. In Exp2, each image belonged to one of three subclasses (air, water, ground). The incongruent prime (vehicle) could belong to the same subclass as the target (related) or not (unrelated). Even though the type of subclass was not a diagnostic attribute for performing the task, related incongruent primes caused more interference than unrelated ones at the longer SOA. Thus, object categories and non-diagnostic attributes are processed automatically. To explain these results, we propose a model based on the idea that categories are represented as hierarchical patterns of neuronal activity in infero-temporal cortex. When two objects share attributes (e.g. 'animal-ness'), their patterns overlap. Since related incongruent primes share more attributes with the target than unrelated ones, their patterns overlap more, increasing the difficulty to determine the target's category. This effect is seen early when shared attributes are diagnostic for the task (e.g. bird primed by dog in Exp1) but later when they are not diagnostic (e.g. bird primed by plane in Exp2). We conclude that visual objects are automatically processed and the residual activity in infero-temporal cortex affects subsequent categorization depending on the extent of overlap between activity patterns.

Introduction

Visual categorization of natural stimuli can be performed extremely rapidly (Crouzet, Kirchner, & Thorpe, 2010; Kirchner & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996) and efficiently (Li, VanRullen, Koch, & Perona, 2002; Poncet, Reddy, & Fabre-Thorpe, 2012; Rousselet, Fabre-Thorpe, & Thorpe, 2002). Ultra-rapid object recognition is probably supported by feed-forward activity (Thorpe & Fabre-Thorpe, 2001) in the ventral stream, particularly in the infero-temporal cortex (ITC) (see DiCarlo, Zoccolan, & Rust, 2012 for a review). Interestingly, the similarity between response-patterns in ITC evoked by different objects correlates with our intuitive category structure (Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte et al., 2008). That is, the patterns of activity elicited by distantly related or

unrelated objects (such as an animate and an inanimate object) are more distinct than the patterns of activity elicited by similar objects (such as two animate objects). Moreover, the difference in ITC activity patterns is also correlated with judgments of similarity between objects (Connolly et al., 2012; Mur et al., 2013). These studies have clear implications for our understanding of the organization and role of ITC in object categorization. The question we would like to address in this study is: how do multiple simultaneously active visual categories in ITC interact? One way to test such interactions between visual categories, if they do occur, is by using a priming paradigm where a first distractor object (prime) is presented just before a second target object. If the activity elicited by a stimulus persists at least for a short while, then its effects on a subsequently activated (and hence concurrently active) representation can be assessed. However, this paradigm has been mostly used with masked primes, that is, when the processing of the prime is interrupted and therefore processed unconsciously (e.g. Dehaene et al., 1998; Dell'Acqua & Grainger, 1999; Van den Bussche, Notebaert, & Reynvoet, 2009). However, we intended to test the interaction between two simultaneously active visual representations (and hence without masking the prime). Furthermore, varying the relationship between prime and target allows us to study the influence of semantic distance between two objects on visual category processing.

Previous studies that examined interaction between objects have primarily done so in the language domain, where various stimuli (pictures, line drawings, words), response modalities (oral versus manual responses), and temporal windows (presentation of the two stimuli with different time intervals between two) were employed. These studies were chiefly interested in determining the interaction between linguistic concepts or between language-based responses. These studies have uncovered, over a couple of decades, a mixed bag of findings, but a coherent story seems to be finally emerging. Picture-word interference paradigms have shown that participants could verbally name a picture faster when the distractor word was identical to the picture (e.g. the word 'cat' with a picture of a cat) than when it was unrelated to it ('table', picture of a cat). In contrast, picture naming was slower when the word was semantically related to the picture ('dog', picture of a cat) than when it was unrelated (e.g. Lupker, 1979; Rosinski, 1977). However, this last interference effect is very sensitive to the stimuli used and the task. For example, it could be completely reversed when target and distractor converge on the same response (Costa, Mahon, Savova, & Caramazza, 2003; Kuipers, La Heij, & Costa, 2006; Kuipers & La Heij, 2012; Lupker & Katz, 1981). A facilitation effect is also observed when the distractor word is a subcategory of the expected response ('seagull', picture of a bird) (Glaser & Dünghoff, 1984; Hantsch, Jescheniak, & Mädebach, 2012) or when it is related but belonged to a different category ('nest', picture of a bird) (Alario, Segui, & Ferrand, 2000; Blackford, Holcomb, Grainger, & Kuperberg, 2012; Costa, Alario, & Caramazza, 2005). Putting all these varied findings together, Mahon, Costa, Peterson, Vargas, & Caramazza, (2007) suggested that the effect of semantic relatedness between target and distractor is influenced primarily by the task: if both the word and the image convey the same response, performance is improved. When the distractor word has the same degree of response convergence as the target, then the more semantically related it is to the image, the faster the responses are. For example, participants are faster in naming an image of a horse when presented with the word 'zebra' than with the word 'whale'. Thus, the effect of a distractor word would

depend on the response criteria determined by the task and by the semantic distance between the distractor and the target when the response criteria are constant.

The interactions between objects in these earlier experiments primarily occurred in the language-based semantic space or at the response decision or production stage. Here, we aimed to examine the interactions within the space of visual categories using a paradigm that is tailored to elicit interactions in visual areas. We tested participants in a manual yes/no categorization task (no involvement of language areas) using a picture-picture priming paradigm. In the visual domain, there is little research that examined interaction between visual object categories. A few studies have tried to use the repetition suppression effect (decrease in activity with the repetition of the same stimulus, Desimone, 1996; Miller, Li, & Desimone, 1991) to find the neural correlates of categorical relationships. For example, EEG and MEG studies reported that the repetition of two identical faces elicited the same adaptation of the N170 component as when two different faces were repeated (Amihai, Deouell, & Bentin, 2011) or when different types of face stimuli (e.g. photograph preceded by line-drawings) were repeated (Harris & Nakayama, 2007). fMRI results are not always consistent but in general, repetition suppression has been found in category selective and non-selective regions of the ventral stream (Chao, Weisberg, & Martin, 2002; Fairhall, Anzellotti, Pajtas, & Caramazza, 2011; Koutstaal et al., 2001; Pourtois, Schwartz, Spiridon, Martuzzi, & Vuilleumier, 2009; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003; Weiner, Sayres, Vinberg, & Grill-Spector, 2010). In distinction to these fMRI studies that attempted to determine whether different areas represent different levels of abstraction along the ventral stream, the purpose of our study was to understand how simultaneously active object categories and their relationship influence object categorization. Surprisingly, to the best of our knowledge, a picture-picture priming paradigm has never been used to test the effect of simultaneously active object representations.

There are good reasons to think that different categories could interact in the ITC when two objects are presented in rapid succession. First, superordinate and basic visual object categories can be accessed automatically and rapidly by the ventral visual areas (Li et al., 2002; Poncet et al., 2012; Rousselet, Thorpe, & Fabre-Thorpe, 2004). Therefore, it seems plausible that the category of a distractor image could influence the processing of a target image, if they are presented within a short temporal window. Second, many studies have reported selective responses of ITC cells as early as 100 ms after stimulus onset (Z. Liu & Richmond, 2000; Perrett, Rolls, & Caan, 1982; Tovée, 1994). Indeed, accurate and robust information about object category can be decoded in ITC from 100 ms after stimulus presentation in monkeys (Hung, Kreiman, Poggio, & DiCarlo, 2005) and humans (H. Liu, Agam, Madsen, & Kreiman, 2009). Such early decoding of category membership has also been shown with EEG and MEG signals with a peak around 120-130 ms (Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; Carlson, Tovar, Alink, & Kriegeskorte, 2013; Murphy et al., 2011; Simanova, van Gerven, Oostenveld, & Hagoort, 2010). Finally, ITC neurons are still active long after the presentation of an unmasked stimulus (Keysers, Xiao, Foldiak, & Perrett, 2005; Rolls & Tovée, 1994). Rolls & Tovée (1994) reported that neurons in ITC responded to a 16 ms stimulus for 200-300 ms. Therefore, when a participant has to categorize a second image presented within 200 ms after a first image, ITC neurons

are still responding to the presentation of the first stimulus. We should then see an effect of the activity induced by a prime picture in the ITC on the categorization of a subsequent one.

In this study, we explored the effect of neuronal activity induced by a distractor image (prime) on the categorization of a subsequent image (target). We first tested participants in a bird/non-bird categorization task. A to-be-ignored prime was presented 80 or 180 ms before the target image. To study the effect of semantic distance between objects, we used three types of prime. They could be from the same category as the target (e.g. another bird), from a related category (e.g. another animal) or from an unrelated category (e.g. a vehicle). From picture word interference and masked priming results, we expect participants to be faster and/or more accurate in the same category condition (congruent trial) than in the two other conditions (incongruent trial). Within the incongruent condition, we might find no effect of semantic distance between the prime and the target categories (no difference between the two incongruent conditions). On the other hand, since ITC is organized in a continuous manner (Kiani et al., 2007; Kriegeskorte et al., 2008) we might be able to observe the influence of this organization on the processing of the target category. In this case, both incongruent conditions should lead to different categorization performance.

Experiment 1: Priming effect in basic-level categorization

The main goal of Experiment 1 was to test the effect of a to-be-ignored prime image on the categorization of a subsequent target image. Subjects performed a bird/non-bird categorization (basic-level categorization) using manual responses (no language production). The time interval between the onset of prime and the target (SOA) could be either 80 or 180 ms. The shortest SOA (80 ms) still allowed participants to correctly perceive the order of the stimuli (Theeuwes & Burg, 2013). The prime could convey the same (congruent trial) or a different (incongruent trial) motor response than the target picture. To further test the effect of semantic distance between prime and target, incongruent primes could be either related (non-bird animal) or unrelated (vehicle) to the target category (bird).

Method

Participants

12 volunteers (8 women, 1 left-handed, mean age 25 ± 0.6) participated in this experiment. All participants had corrected or corrected to normal vision and provided written informed consent.

Design

The experiment consisted of a total of 16 blocks of 96 trials each. The target-prime SOA was 80 ms in half of the blocks and 180 ms in the other half. The order of the blocks was randomized. Within each block, half of the target images were bird, $\frac{1}{4}$ were non-bird animals and $\frac{1}{4}$ were vehicles (chance level

was at 50%). As target images included other animals, the bird/non-bird categorization had to be performed at the basic-level (not at a superordinate-level, see Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009). For each target image, half of the primes were congruent (both prime and target conveyed the same motor response), and half were incongruent (prime and target conveyed different motor responses). Moreover, incongruent primes were divided in two equal groups. They were either related (non-bird animal) or unrelated (vehicle) to the target category. The type of trial was intermixed within each block.

Procedure

Participants were seated in a dimly lit room, approximately 60 cm from a CRT screen (1024x768 pixels, refresh rate: 100 Hz). Stimuli were displayed using Matlab with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) and were synchronized with the refresh rate of the monitor. Participants were asked to perform a bird/non-bird categorization task. At the beginning of a trial, a white fixation dot was presented at the center of a black screen for 500 ms. The prime image was then presented at the center of the screen for 20 ms followed by a black screen. The target image presented for 100 ms appeared either 80 or 180 ms after the onset of the prime (Figure 1). Participants were told to ignore the first image and report whether the second image contained a bird or not by pressing the left ('yes' response) or right ('no' response) arrow key using the same hand. The next trial started right after the participant's response or after a maximum of 1.2 sec after the onset of the target image presentation. At the end of the trial, an auditory beep (negative feedback) signaled an incorrect response or an absence of response. Before the beginning of the experiment, participants were trained on 20 trials with a separate set of images.

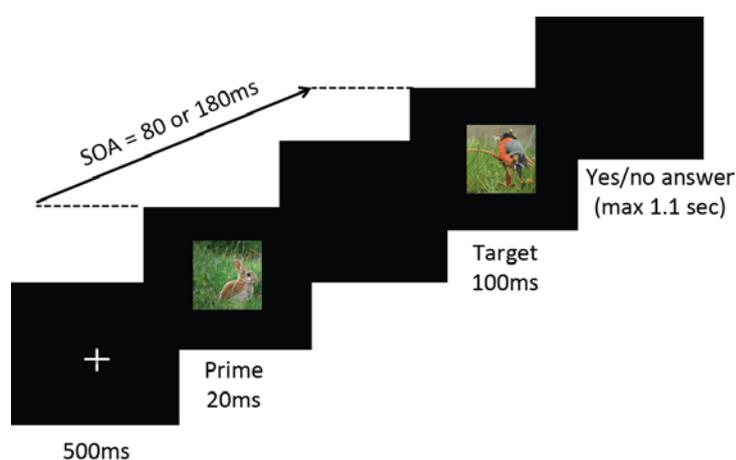


Figure 1. Illustration of a trial in Experiment 1 and 2. In Experiment 1 participants were asked to perform a bird/non-bird categorization on the target image whereas in Experiment 2 they were asked to perform an animal/non-animal categorization. In both experiments participants were told to ignore the prime image.

Stimuli

All 3072 images used in the experiment were colored natural images chosen from the image sets used in previous studies (Poncet & Fabre-Thorpe, 2014) and from the Internet. They contained one or

several objects of various orientations, positions and sizes. The bird category (1536 images) included birds of different species (songbirds, owls, parrots, pigeons, ducks, etc.) in different environments (flying in the sky, perched on a tree, on the water, etc.). The non-bird animal category (768 images) included mammals (dog, bear, elephant, etc.; 384 images) and fish (redfish, shark, etc.; 384 images). The vehicle category (768 images) was also diverse and included planes, cars, boats etc. Each image subtending 8°x 8° visual angle was seen only once by each participant either as a prime or as a target (its status was randomly assigned).

Data analysis

For each participant, behavior was assessed using median reaction time (RT) for correct yes responses (hits) and mean error rates (ER) on bird target images (misses). Anticipatory responses (RT faster than 150 ms) and trials without any response (no response within the 1.1sec) were rejected from the analysis (0.5%).

As a first step, we performed an omnibus 2 (target-prime congruency: congruent and incongruent) x 2 (SOA: 80 and 180 ms) repeated measures ANOVA on raw RTs and ERs. Then, to study the effect of semantic distance on the interaction between categories, we analyzed the effect of the similarity between prime and target categories. To do so, we first defined the priming effect as *the difference in performance (RT or ER) between incongruent and congruent trials*. Then, we calculated priming effects for the two target-prime relatedness conditions separately: where the incongruent prime was either related (non-bird animal) or unrelated (vehicle) to the target (bird). Since we used the same congruent condition (bird prime followed by a bird target) in both cases, differences in the priming effect reflected differences between the two incongruent conditions. We performed a 2 x 2 repeated measures ANOVA (prime relatedness x SOA) on the priming effect for RTs and ERs (henceforth referred to as RT and ER priming). When necessary, further paired t-tests were conducted with appropriate Bonferroni corrections applied for multiple comparisons.

Results

Congruency effect:

Participants were 36±5 ms faster ($F(1,11)=69.49$, $p=4 \times 10^{-06}$, $\eta^2=0.86$) and 3±1% more accurate ($F(1,11)=15.31$, $p=0.002$, $\eta^2=0.58$) in responding to a bird after seeing a bird prime than after seeing an incongruent prime (Figure 2A). There was no effect of SOA on RTs (no main effect $F(1,11)=0.02$, $p=0.90$ or interaction $F(1,11)=2.16$, $p=0.17$) or on ERs (no main effect $F(1,11)=1.25$, $p=0.29$ or interaction $F(1,11)=0.34$, $p=0.57$). The difference between congruent and incongruent trials suggests that even though participants knew that they had to ignore the first (prime) image, they automatically processed its basic category, which then affected their categorization responses to the target image.

Influence of prime relatedness on priming effect:

ER priming was larger for non-bird animal primes than for vehicle primes ($F(1,11)=9.21$, $p=0.011$, $\eta^2=0.46$) (Figure 2B). There was no effect of SOA ($F(1,11)=0.36$, $p=0.56$) or interaction ($F(1,11)=1.90$, $p=0.19$). That is, participants made more mistakes when incongruent primes belonged to the same superordinate category (animal) as the target (bird) than when they belonged to a different one (vehicle).

Although there was no main effect of prime relatedness ($F(1,11)=2.90$, $p=0.12$) or SOA ($F(1,11)=3.00$, $p=0.11$) on RT priming, we found a significant interaction between these two factors ($F(1,11)=8.74$, $p=0.013$, $\eta^2=0.44$). At 80 ms SOA, RT priming was the same for non-bird animal and vehicle primes (paired t-test $t(11)=1.12$, $p=0.6$). In other words, the type of incongruent prime (related or unrelated) had no influence on RTs. At 180 ms, RT priming was larger for non-bird animal than for vehicle primes (paired t-test $t(11)=3.41$, $p=0.01$). Thus, at longer SOA, incongruent related primes (animals) slowed down the categorization of the target picture (bird) more than incongruent unrelated primes (vehicles).

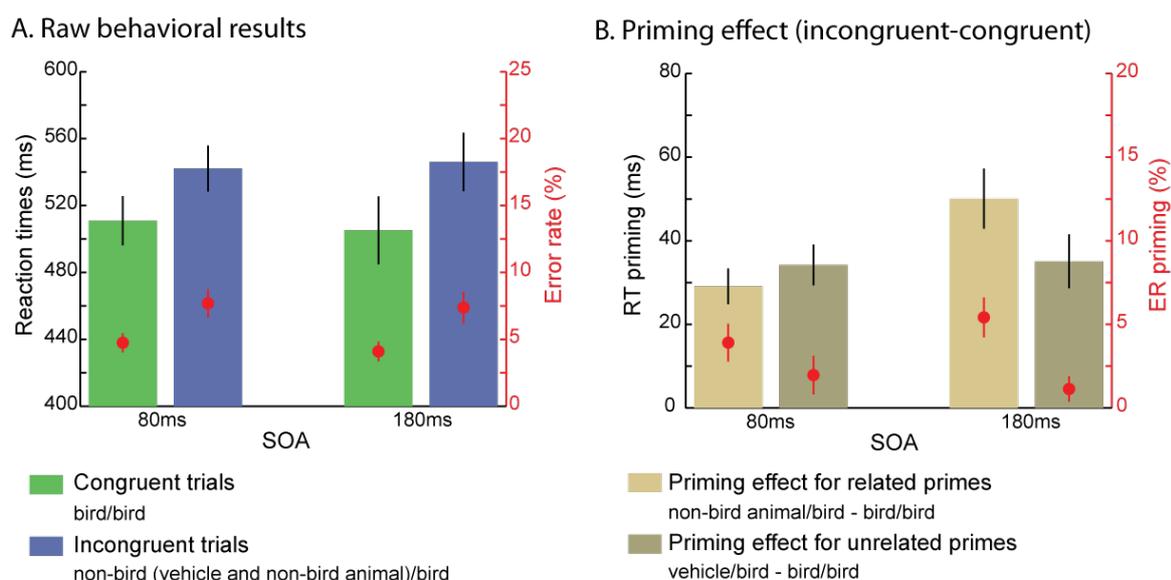


Figure 2. Experiment 1: Performance in bird/non-bird categorization task. **Panel A: Congruency effect.** Bird targets were categorized faster (bars) and more accurately (red circles, see right vertical-axis) when they were preceded by a congruent prime (bird, green bars) than an incongruent prime (non-bird, blue bars) at both 80 ms and 180 ms SOAs. **Panel B: Effect of prime relatedness on the priming effect (difference between incongruent and congruent primes).** At both SOAs the ER priming was larger for related (non-bird animal, red circles within light brown bars) than for unrelated (vehicle, red circles within dark brown bars) primes. RT priming (bars) was the same for related and unrelated primes at short SOA. However, this RT priming was higher for related compared to unrelated primes at long SOA.

Discussion

Our results showed that an image flashed for only 20 ms affected the categorization of a subsequent image both in terms of response speed and accuracy. If both images in the sequence contained a bird

(even though not identical), participants were around 36 ms faster and more accurate (3% less errors) than when the prime was a non-bird animal or a vehicle. This was true at short and long SOA implying that some information about the prime, perhaps something sufficient for a motor response, had been processed very quickly.

As such, it can be argued that this priming effect is not the consequence of the interaction between different object categories but reflects stimulus-response association for the decision or a conflict between different motor responses (Damian, 2001; Eckstein & Henson, 2012; Kiesel, Kunde, & Hoffmann, 2007). Indeed, decrease in reaction times to categorize bird images as bird can be explained by the repetition of the same task on the same basic category stimuli (Denkinger & Koutstaal, 2009; Horner & Henson, 2009). However this possibility cannot explain why we observed different priming effect depending on the category of the incongruent prime. Non-bird images were from different basic categories, so stimulus-response associations were not straightforward to build. More importantly, both type of incongruent primes were presented the same number of times over the experiment and have the same stimulus-response association (“no” response). Thus, if the priming effect was only driven by stimulus-response binding or a motor conflict, then we should observe the same amount of priming for related and unrelated primes. This suggests that at least a part of the priming effect observed, if not all of it, is directly caused by the similarity between the prime and the target category.

Specifically, our results showed that the priming effect was larger for related (non-bird animal) than unrelated (vehicle) primes and increased with longer SOA for related primes only. These results appear at odds with the ones reported in picture word interference paradigms. For example, in a study by Kuipers & La Heij (2012), participants’ performance on a word categorization task was the same whether the distractor picture, different from the target category, was semantically related or unrelated to the word. However, the current results might be different because our study minimizes the involvement of language areas (either for reading the target word or producing verbally the answer). Unlike in that report, we argue that the observed interaction occurred over purely visual representations. Thus in pure visual categorization, prime relatedness affects performance. The activity evoked by a similar representation (non-bird animal) interferes more with the processing of a target image (bird) than the activity evoked by a less similar representation (vehicle). We propose below a biologically plausible model to account for our results and explain them in terms of competition between different patterns of neuronal activity.

A physiologically inspired model

Our model is based on three observations and inferences made in the physiological literature.

1. *Categorical representation is hierarchically organized in the visual cortex.* Category information about visual objects is most probably represented in ITC (DiCarlo et al., 2012). Moreover, the pattern of ITC neuronal activity evoked by different object categories appears to be represented hierarchically

(Kiani et al., 2007; Kriegeskorte et al., 2008): similar objects (e.g. a dog and a bird) evoke a more similar pattern of activity than less similar objects (e.g. a car and a bird). Thus, we assume that the activity evoked by a stimulus is represented by a co-activation of neurons coding for various attributes organized hierarchically (Figure 3A).

2. *Activity invoked in the ITC by images takes time to decay.* ITC neurons are active long after the presentation of an unmasked stimulus (Keysers et al., 2005; Rolls & Tovée, 1994). There is also ample evidence to support the idea that objects are represented by neuronal populations collectively firing in the gamma frequency (~40 Hz) range (see Fries, 2009; Jerbi et al., 2009; Lisman & Jensen, 2013; Roux & Uhlhaas, 2014; Singer & Kreiman, 2009; for reviews) and that these gamma bursts last for a long time, over a period of 300 – 600 ms, even if the stimulus is physically removed (Jensen, Kaiser, & Lachaux, 2007; Vidal et al., 2010). In other words, if two stimuli are presented in rapid succession, the induced activations, whatever their nature, should interact and summate. We assume a linear summation of signals, as that is the simplest hypothesis. However, a non-linear summation would not alter the conclusions of our analysis. In our paradigm, two natural images are presented in rapid succession (at SOAs of 80 and 180 ms in two separate conditions), well within the decay period of the activity induced by a visual stimulus. Therefore, we suggest that the resulting activity in ITC is a linear summation of the activity from prime and target images (Figure 3B).

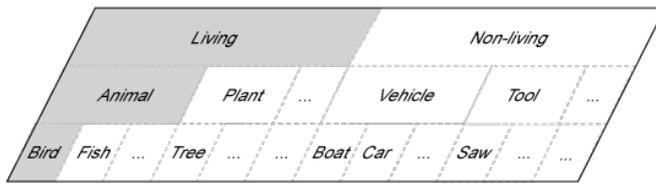
3. *Cortical circuits ‘read’ off the code in the ITC for making decisions.* This assumption is not controversial (DiCarlo et al., 2012). Some area of the brain has to convert the code present in the ITC into motor responses (verbal or manual). This read out can be performed by a separate part of ITC itself or the Pre-Frontal Cortex (PFC; for e.g., see Freedman, Riesenhuber, Poggio, & Miller, 2001, 2003; Roy, Buschman, & Miller, 2014). This decision-making is straightforward when a single categorical (or unambiguous) representation is active, as in the congruent condition. However, when the ITC activity does not convey information about a single representation, but a summated signals of two distinct categories, as in the incongruent condition, the decision-making areas have to engage in a process analogous to signal detection analysis to separate out the two summated representations. We propose that if the prime and target images are closely related, then this ‘signal detection’ analysis would be harder and take longer, whereas if the signal is a summation of very dissimilar categories, the analysis should be easier and faster. Indeed, there is evidence that categorical decisions take longer if an object is ambiguous compare to when it is unambiguous (Roy et al., 2014).

With these observations in place, we aim to explain the various results we observed in this experiment. The overall priming effect (facilitation in congruent trial and interference in incongruent trial) can be easily explained by the fact that determining the category of a single representation (congruent condition) should take less time than determining the category of two summated representations (incongruent condition). Further, in a congruent trial both the prime and the target induce the same motor response whereas in an incongruent trial they induce opposite responses. Both reasons should speed up congruent responses and impede incongruent responses.

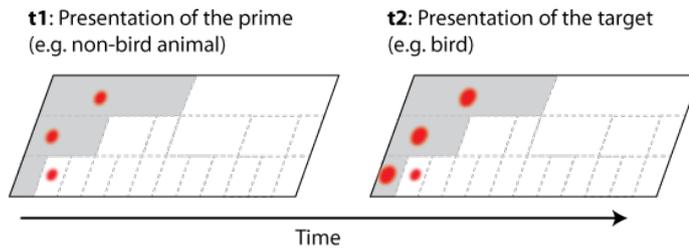
What is more interesting is the difference between the impact of related and unrelated incongruent primes on target categorization. As noted above, when a non-bird animal is presented before a bird, some attributes are activated by both stimuli (e.g. 'being an animal') such that the pattern of activity of the prime and the target overlap (Figure 3C, incongruent related trial). In the case of a vehicle prime, there are no shared attributes between the two stimuli so the two patterns of activity are easily distinguishable (Figure 3C, incongruent unrelated trial). Hence, the decision-making areas will take longer to determine the target category when it is preceded by a related prime than by an unrelated prime, as found in our study (Fig 2B, left bars). A different but qualitatively similar description is that at the end of an incongruent trial, there is a competition in the decision-making areas between prime and target representations. The winner of this competition between representations will access the appropriate response ('bird' or 'not-bird'). This competition is more intense and less easy to resolve when the two objects share some attributes than when they do not.

We also argue that the pattern of neuronal activity changes over time (as shown in neurophysiological studies, e.g. Meyers, Freedman, Kreiman, Miller, & Poggio, 2008). At 180 ms SOA the activity in ITC has been sustained for a longer time and more information about the prime will have been accessed (increase of activity and/or recruitment of new neurons and/or inhibition). When considering an unrelated prime (vehicle) this additional information will not have any influence on the priming effect, as the summated signal still consists of vastly different category attributes and the decision-making areas should have no difficulty distinguishing the two representations. However, in the case of a related prime (non-bird animal), the additional information activates attributes that are common with those of the target image (e.g. living object). Thus, the activity pattern evoked by a related prime becomes more similar to the one of the target at long SOA than at short SOA. This would make the decision-making process even more difficult and longer than at 80 ms SOA. That is, it will lead to a stronger competition between the activity evoked by the incongruent related prime and the target predicting a larger priming effect (Fig 2B, right bars).

A. Schematic organization



B. Example of an incongruent trial activity



C. Activity at t2

		Congruent trial	Incongruent trial	Competition (incongruent trial)
80ms SOA	related			++
	unrelated			+
180ms SOA	related			+++
	unrelated			+

Figure 3. Model of Experiment 1 results. **Panel A:** Simplified spatial organization of some categories used in the experiment. Basic-level representations such as bird and car belong to the superordinate categories animal or vehicle which are either living or non-living objects. In a bird/non-bird categorization task, the system would recognize a bird image by a pattern of activity including bird and bird-related (e.g. animal and living object) attributes (grey areas). **Panel B:** Representation of patterns of neuronal activity (in red) elicited during one trial (example of a related incongruent trial). t1: At the presentation of a non-bird animal prime, the population coding for animal attributes quickly reacts (represented in red) to the stimulus as well as some populations coding for more specific basic attributes (e.g. fish). t2: The activity (in red) evoked by a bird target is added to the prime activity. At the end of a trial, a competition takes place between the activity expected for a bird image (grey areas) leading to a “yes” response in the categorization task and the activity evoked by the non-bird prime leading to a “no” response. **Panel C:** Patterns of activity at t2 (end of a trial) in incongruent and congruent trials in different condition (2 SOAs and 2 relatedness). The strength of competition between the activity evoked by the incongruent prime and the one evoked by the target is represented in the right column. It increases with the increase of similarity between the patterns of activity evoked by the two stimuli.

Experiment 2: Priming effect in superordinate-level categorization

In Experiment 2, we tested participants' performance in a similar paradigm as in Experiment 1 but using a superordinate categorization task (animal/non-animal). The model developed to explain the results of Experiment 1 predicts that the amount of priming (incongruent – congruent performance) depends on the representational similarity (or shared attributes) between the incongruent prime and the target: the higher the similarity between the two, the stronger the priming effect. In Experiment 1 the shared attributes between prime and target that caused competition between them, and hence resulted in priming, were diagnostic for the task to be performed. Both a related prime (say, dog) and the target (bird) shared the property of 'animal-ness' and presumably both evoked the parts of the representation expected for the target category (grey areas in Figure 3C). 'Animal-ness' is an inherent property of being a bird or a dog and activating this property is essential for identifying their category. To further push our model, in Experiment 2 we tested whether non-diagnostic features could also influence the priming effect by using attributes that are not inherent to a specific superordinate category. For example, we tested whether an airplane ('air object') prime would interfere more than a car ('ground object') prime when the target animal is a bird ('air animal'). In this situation, the shared attribute 'air-ness' is not a relevant property for the task (animal/non-animal discrimination).

Method

Participants

12 new volunteers (7 women, no left-handed, mean age 24 ± 0.8) were included in this experiment. An additional participant was excluded from analysis, as he did not reach the accuracy criterion (at least 80% accuracy in each condition). All participants had corrected or corrected to normal vision and provided written informed consent.

Design

The experiment consisted of a total of 16 blocks of 96 trials. In half of the blocks the SOA between the prime and the target image was 80 ms and in the other half it was 180 ms. The order of the blocks was randomized. Within a block, half of the target images contained an animal and half a vehicle. For each target types, half of the primes were congruent and half were incongruent. Further, each stimulus could belong to one of three subclasses (air, ground or water, see the section on Stimuli). For each congruent or incongruent condition, half of the primes belonged to the same subclass as the target image (related prime) and half belonged to another subclass (unrelated prime). Thus in one block, participants performed 14 congruent related trials, 14 congruent unrelated trials, 14 incongruent related trials and 14 incongruent unrelated trials with an animal presented as the target image. The same conditions with the same number of trials per block were used but with a vehicle presented as the target image.

Procedure

The same procedure as in Experiment 1 was used but participants were asked to perform an animal/non-animal categorization task (Figure 1).

Stimuli

Most of the stimuli used in this experiment were the ones used in Experiment 1 but for the purpose of the experiment, fewer images of birds and more images of vehicles and non-bird animals were included. The set of images was composed of 1536 animals and 1536 vehicles that could be presented as either the prime or the target image. These 2 categories were composed of 3 equal subclasses: air, ground and water. In the animal category, 1/3 of the images were birds (duck, pigeon, etc.), 1/3 were mammals (cow, deer, etc.) and 1/3 were fish (shark, salmon, etc.). In the vehicle category, 1/3 of the images were ground vehicles (motorbike, truck, etc.), 1/3 were air vehicles (plane, helicopter, etc.), and 1/3 were water vehicles (boat, canoe, etc.). Within each subclass, the object was presented in various contexts, for example a bird could be standing on the ground or on a lake, flying in the sky, etc.

Data analysis

As in Experiment 1, we calculated median RT and ER on animal target pictures for each participant. 0.4% of all trials (anticipatory responses or no response) were rejected from the analysis.

First, we conducted a 2x2 repeated measures ANOVA (congruency x SOA) on raw RT and ER regardless of prime relatedness to look at the effect of prime-target congruency. As in experiment 1, we defined the priming effect as the difference in performance in incongruent and congruent trials. We determined the priming effect for related images (e.g. plane/bird trial – bird/bird trial) and for unrelated images (e.g. plane/mammal – fish/mammal), separately. A 2x2 repeated measures ANOVA (SOA x relatedness) was applied to these RT and ER priming effects. As in Experiment 1, when required, paired t-tests were applied on the results with Bonferroni correction.

Results

Congruency effect:

Participants were 47 ± 7 ms faster ($F(1,11)=58.61$, $p=1 \times 10^{-05}$, $\eta^2=0.84$) and $3.5 \pm 1\%$ more accurate ($F(1,11)=19.58$, $p=0.001$, $\eta^2=0.64$) to categorize animal images when they were preceded by an animal prime than when they were preceded by a vehicle prime (Figure 4A). Although we found no main effect of SOA on ERs ($F(1,11)=0.13$, $p=0.72$) or RTs ($F(1,11)=0.38$, $p=0.55$), we observed a significant interaction for RTs ($F(1,11)=8.38$, $p=0.015$, $\eta^2=0.43$). Further analysis revealed that the RT priming effect was larger at 180 ms SOA (54 ± 7 ms) than at 80 ms SOA (40 ± 6 ms) (paired t-test $t(11)=2.89$, $p=0.01$). As found in Experiment 1, the prime category was processed automatically by the participants leading to better performance in the congruent condition than in the incongruent condition.

Influence of prime relatedness on priming effect:

The priming effect (incongruent-congruent, Figure 4B) on ERs was similar at both SOA ($F(1,11)=0.77$, $p=0.40$) and for both related and unrelated primes ($F(1,11)=0.70$, $p=0.42$; no interaction $F(1,11)=0.06$, $p=0.81$). Considering the priming effect on RTs, we found no main effect of prime relatedness ($F(1,11)=1.60$, $p=0.23$) but a significant effect of SOA ($F(1,11)=9.03$, $p=0.01$, $\eta^2=0.4$) and interaction between these two factors ($F(1,11)=7.40$, $p=0.02$, $\eta^2=0.40$). RT priming increased significantly at longer SOA for related primes (from 36 ± 8 ms to 61 ± 7 ms, paired t-test, $t(11)=3.52$, $p=0.01$) but not for unrelated primes (from 38 ± 7 ms to 47 ± 8 ms, paired t-test, $t(11)=1.60$, $p=0.28$).

Because of the design used in Experiment 2, the increase in RT priming for related primes could be explained by either a) a decrease of RTs for congruent trials, b) an increase of RTs for incongruent trials or c) both. To test it, we compared RTs at 80 and 180 ms for congruent related and incongruent related trials separately. This analysis showed a marginal difference for both congruent (paired t-test, $t(11)=1.98$, $p=0.15$; $p=0.07$ before Bonferroni correction) and incongruent (paired t-test, $t(11)=2.18$, $p=0.10$; $p=0.052$ before Bonferroni correction) trials. Thus, the increase of RT priming for related trials at long SOA is a combined effect of a RT decrease in the congruent condition (-7 ± 4 ms) and a RT increase in the incongruent condition ($+16\pm 8$ ms).

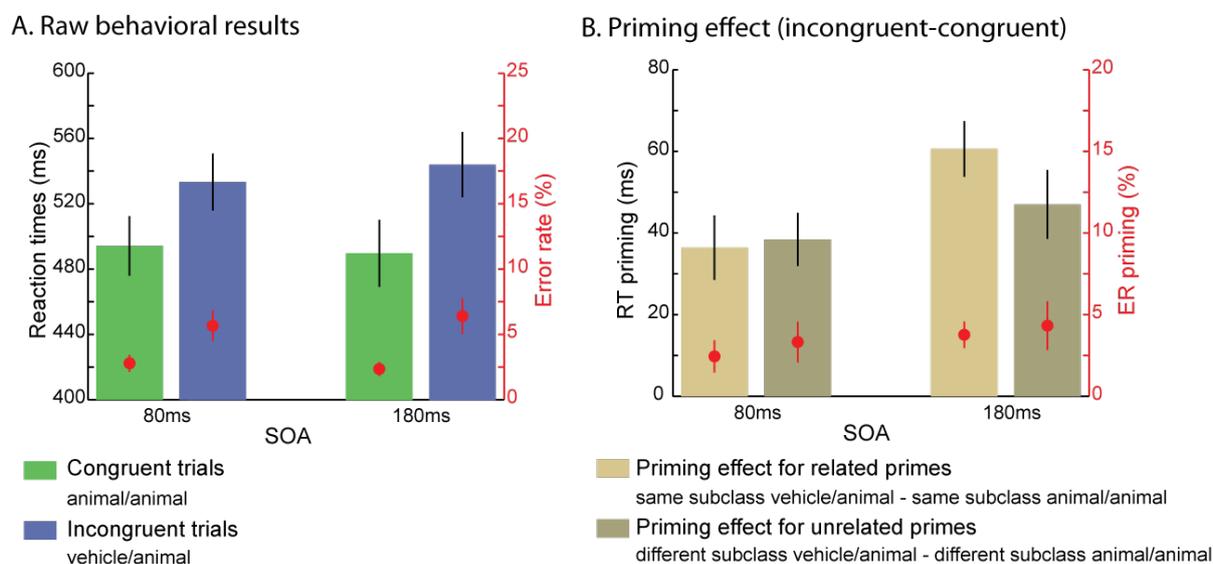


Figure 4. Experiment 2: Performance in animal/non-animal categorization task. Each image (prime and target) belonged to one of three subclasses (air, water, ground). Legend as in Figure 2. **Panel A: Priming effect.** Performance was better (shorter RTs and lower ERs) in the congruent than in the incongruent condition. **Panel B: Effect of prime relatedness on the priming effect.** Prime and target were either related (same subclass, e.g. both air objects) or unrelated (different subclasses). There was no effect of relatedness at 80 ms SOA on the priming effect (incongruent-congruent for each related or unrelated condition separately). At longer SOA, RT priming increased significantly for related prime but not for unrelated primes.

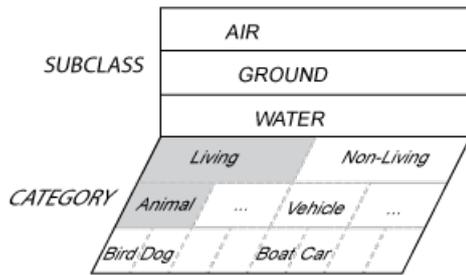
Discussion

Our results clearly showed that a briefly presented image influenced the categorization of a following image in an animal/non-animal task. Despite a large diversity of animals (and vehicles) used in this study, we found a large priming effect. Participants were slower (by around 47 ms) and less accurate (around 3.5% increase in error rate) in categorizing an animal target after a vehicle prime than after an animal prime. Subclass relatedness had no influence on the priming effect at 80 ms SOA. However, the priming effect increased at 180 ms SOA for related primes but not for unrelated primes. Thus, even though the two objects did not share any common category (one was a living object and the other a non-living object), we observed an effect of prime relatedness. For example, participants were faster to categorize a ground animal (e.g. a dog) after a ground vehicle (e.g. a car), than after an air vehicle (e.g. an airplane). This result is surprising in the sense that participants were asked not to process the prime (they could have suppressed its processing), let alone additional information that is not required for the categorization task. Yet at 180 ms SOA, the amount of priming depended on how much overlapping information (relevant to the task or not) there was between the prime and target. This confirms that when object representations become more similar, we observe a stronger priming effect.

One possible interpretation of the increase of priming effect for relevant primes is that it is a consequence of image context. For example, air objects such as birds and planes have a sky background whereas water objects are present in (or on) water. Even though we cannot completely exclude this possibility, the images in the database used in this experiment all had various environments. For example, a bird or a plane could be in the sky, on the ground, or on water. Furthermore, there could be other possible explanations of the relatedness effect such as shape similarity (a flying bird and a plane; but all birds are not flying), functionality (both bird and plane are flying objects), among other such similarities. Further experiments are needed to know exactly what attribute(s) is/are at the origin of this effect. In any case, we observed an effect of prime relatedness even though the two objects belonged to different superordinate categories.

The effect of prime relatedness is only visible at 180 ms SOA, implying that non-diagnostic attributes are only processed in a later time window than diagnostic attributes (the effect was already present at 80 ms SOA in Experiment 1). These results provide supporting evidence for our model and further insights about the temporal dynamics of category interactions (Figure 5). At 80 ms SOA, only diagnostic attributes have enough time to be processed leading to a lack of any difference in priming effect between related and unrelated incongruent primes. At longer SOA other attributes that are not needed for the task are also processed automatically. If these non-diagnostic attributes are shared between the prime and the target then they contribute to strengthen the priming effect by virtue of increased similarity between the two representations. Therefore, a larger priming effect is observed at 180 ms SOA.

A. Schematic organization



B. Activity at t2

		Congruent	Incongruent	Competition (incongruent trial)
80ms SOA	related			+
	unrelated			+
180ms SOA	related			++
	unrelated			+

Figure 5. Model of Experiment 2 results. **Panel A:** Simplified spatial organization of some categories used in Experiment 1 and the three subclasses used in Experiment 2 (air, ground, water). In an animal/non-animal categorization task, the system would recognize an animal image by a specific pattern of activity (grey areas). **Panel B:** Patterns of activity (in red) at t2 (end of a trial) in incongruent and congruent trials in different condition (two SOAs and two relatedness). As proposed earlier, at the end of a trial a competition takes place between the patterns of activity expected for an animal image (grey areas) inducing a “yes” response in the task and the activity evoked by the vehicle prime inducing a “no” response. The more similar the two patterns of activity are, the harder the competition (larger priming effect). At 180 ms SOA, more information about the prime is processed and if some attributes are shared with the target (even if they are cross-category attributes) then the activity pattern of the prime becomes more similar to the one of the target. Consequently, the competition between prime and target representations increases.

General discussion

The main goal of this study was to test how visual categories interact in a short temporal window. Our results showed that a brief presentation of an image influenced categorization performance of a subsequent test image at the basic and superordinate-levels. When the two images conveyed the same response ('yes') performance was better (faster RTs and better accuracy) than when the two images conveyed conflicting responses. Participants clearly knew that they should ignore the prime but its processing was unavoidable. Interestingly, we also found that the more the prime was related to the target category (either because they shared superordinate attributes or cross-category task-irrelevant attributes), the larger the priming effect (the larger the performance difference between congruent and incongruent trials). Moreover, this effect was influenced by the time available to process the prime: the longer the SOA the larger the effect of prime relatedness.

To avoid any potential confounds, first, we ensured that participants were not confusing the order of the prime and target image by using long enough SOAs (Theeuwes & Burg, 2013). Second, contrary to usual practice in priming studies, we used each image only once in an experiment. Further, the set of images was chosen so that it included very diverse objects (various sizes, orientations, etc.) so our results cannot be explained by simple image learning (low or high level) or familiarity due to image repetition. Furthermore, the stimuli were picked randomly for each participant, so the effect found in our study cannot be driven by a specific pair of stimuli. Finally, participants were performing only one type of categorization (Experiment 1 participants were different from Experiment 2 participants). Therefore, the association stimulus-response was the same over the experiment (Horner & Henson, 2009, 2011a, 2011b) and any kind of task-dependent interaction between categories was avoided (Evans, Horowitz, & Wolfe, 2011).

We propose a model to explain our results based on observations and inferences from other studies. We argue that the priming effect could be explained by (a) a faster read-out of only one representation in the case of a congruent trial versus two in an incongruent trial and (b) a response congruency effect. Our results also showed an effect of prime relatedness on the priming effect that could be well explained by ITC organization. Studies have shown that the pattern of activity evoked by a bird image is more similar to the one evoked by another animal than by a vehicle (Kiani et al., 2007; Kriegeskorte et al., 2008). We argue that when the system has to make a choice between two different representations (incongruent condition), the speed with which one representation is excluded would increase with the decrease in overlap between the two evoked patterns of activity. Thus, related primes would induce a larger priming effect than unrelated primes. We also argue that patterns of activity are dynamic. Additional information is processed at longer SOA allowing the possibility for further prime attributes (even if irrelevant to the task, Experiment 2) to be processed. If these additional attributes are shared with the target, then the prime's pattern of activity becomes more similar to that of the target, leading to a larger priming effect.

A similar hypothesis of response exclusion has been put forward by Mahon et al., (2007) to explain semantic interference in language production (see also Finkbeiner & Caramazza, 2006). They argue

that an unrelated distractor word could be excluded faster than a distractor word that satisfied a response criterion required by the target. Our results showed a similar effect in Experiment 1: the amount of interference on bird target images was larger following non-bird animal than vehicle prime images. The non-bird animal prime satisfied a response criterion required by the task (being an animal), which is not the case for a vehicle prime. Mahon et al., (2007) also showed that when the response criterion is constant, related distractors lead to shorter naming latencies than unrelated ones. However in Experiment 2, incongruent primes were from the same category (vehicle) and our results showed that the interference was higher after a related prime than an unrelated one but needed more time to develop as it was only observed at 180 ms SOA. Thus our results do not reflect the same mechanisms observed in studies involving language and we argue that the priming effect observed here reflects a competition in the ventral stream between visual representations elicited by the prime and the target, and not just competition at the response level.

We assumed that the interaction between prime and target would take place in ITC where category representations arise. However, it is possible that these interactions take place at a later step, in the prefrontal cortex (PFC) or even in the lateral or medial intraparietal areas (Fitzgerald, Freedman, & Assad, 2011; Swaminathan & Freedman, 2012; Swaminathan, Masse, & Freedman, 2013). ITC has strong projections to PFC (Ungerleider, Gaffan, & Pelak, 1989; Webster, Bachevalier, & Ungerleider, 1994) and it has been argued that while ITC encodes relevant information needed to identify an object category, PFC categorizes stimuli based on even more abstract attributes (Freedman et al., 2003; Pan & Sakagami, 2012). On the other hand, in priming studies PFC seems to be mostly involved in the retrieval of stimulus-response associations (Horner & Henson, 2008, 2011a; Thompson-Schill, D'Esposito, & Kan, 1999) and such associations cannot explain the differences we observed between related and unrelated primes that induce the same "no" response. Moreover, it has been shown that monkeys with bilateral removal of PFC were still able to learn and generalize perceptual categories of related stimuli rapidly without explicit instruction (Minamimoto, Saunders, & Richmond, 2010). In humans, damage to ITC but not PFC leads to visual categories agnosias (Gainotti, 2000). These results as well as the extremely rapid reaction times observed in humans when performing categorization tasks (e.g. Thorpe et al., 1996), lead to the conclusion that PFC might not be involved in easy categorization tasks. However, PFC might play a more important role in the priming effect observed in Experiment 2 at 180 ms SOA. In this case, categories were perceptually dissimilar but we observed an effect of relatedness which could be the result of a more conceptual processing that involves PFC.

Our model is not only explaining the data that we observed in our study but can also account for findings in rapid visual categorization tasks. Many studies have reported that performance (RTs and ERs) was lower in basic-level categorization tasks when targets and distractors shared the same superordinate category (e.g. dogs vs. cats categorization) than when they were not (e.g. dogs vs. buses categorization) (Bowers & Jones, 2008; Macé et al., 2009). In fact, Macé et al. (2009) showed that in a bird/non-bird categorization task, participants made more errors for objects similar to the target (insects) than for the ones that were more different (e.g. mammals). Moreover, Mohan & Arun (2012) have shown that the time to categorize an object could be predicted by its similarity with other objects.

From the perspective of our model, the pattern of activity from an animate object competes less with (is less similar to) a pattern of an inanimate object than with a pattern of another animate object. Therefore, we would expect performance to be higher when the categorization task involves distant categories (superordinate-level) than when it involves closer categories (basic-level). This is indeed what has been found in recent studies reporting a superordinate advantage in object (Macé et al., 2009; Poncet & Fabre-Thorpe, 2014; Poncet et al., 2012; Praß, Grimsen, König, & Fahle, 2013) and scene categorization (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Kadar & Ben-Shahar, 2012; Loschky & Larson, 2010).

We acknowledge that even though our model is based on findings from electrophysiological studies and explains our findings as well as those from rapid categorization studies, it still needs to be tested further. This could be done by recording neuronal population but a simpler way to test our model could be by using morphed stimuli. For example, one could test the effect of a morphed prime between a bird and a dog in a similar protocol as the one used in Experiment 1. Our model predicts that when the morphed prime is predominantly composed of bird attributes, and therefore categorized as a bird, a following bird image should be processed as if a bird prime was presented. On the other hand, if a morph prime is predominantly composed of dog attributes, and is categorized as a dog, then the prime will compete with the target representation (leading to longer response latencies and/or higher errors). On the contrary, if the interference that we observed in our study is caused by low level features and not by categorical representations (that is before IT), then we would observe a linear increase in reaction times and/or error rate with the increase of prime morphing towards the incongruent category.

We think that the method used here as well as the results we obtained are very encouraging for future research on visual category organization. Indeed our protocol does not use resource-intensive techniques, such that any new hypothesis is easy to implement and test. Furthermore, as shown in Experiment 2, this protocol could be used not only to test visual representations but also the formation of more conceptual representations. Specifically in this study, by testing how categories interact, we have shown some evidence in favor of a competition between visual representations: the more similar the representations, the higher the competition.

References

- Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 53(3), 741-764. doi:10.1080/713755907
- Amihai, I., Deouell, L. Y., & Bentin, S. (2011). Neural adaptation is related to face repetition irrespective of identity: a reappraisal of the N170 effect. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, 209(2), 193-204. doi:10.1007/s00221-011-2546-x
- Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, 123(1), 84-99. doi:10.1016/j.cognition.2011.12.007
- Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology (2006)*, 61(4), 552-557. doi:10.1080/17470210701798290
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10(4), 433-436.
- Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, 11(10), 9-9. doi:10.1167/11.10.9
- Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, 13(10). doi:10.1167/13.10.1
- Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent Modulation of Category-related Cortical Activity. *Cerebral Cortex*, 12(5), 545-551. doi:10.1093/cercor/12.5.545
- Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(8), 2608-2618. doi:10.1523/JNEUROSCI.5547-11.2012
- Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, 12(1), 125-131.
- Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorisation effect: A novel effect in the picture-word interference paradigm. *Language and Cognitive Processes*, 18(2), 205-234. doi:10.1080/01690960143000524
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision*, 10(4), 16.1-17. doi:10.1167/10.4.16
- Damian, M. F. (2001). Congruity effects evoked by subliminally presented primes: automaticity rather than semantic processing. *Journal of Experimental Psychology. Human Perception and Performance*, 27(1), 154-165.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., ... Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, 395(6702), 597-600.

- Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, 73(1), B1-B15.
- Denkinger, B., & Koutstaal, W. (2009). Perceive-decide-act, perceive-decide-act: How abstract is repetition-related decision learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 742-756. doi:10.1037/a0015263
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, 93(24), 13494-13499.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415-434. doi:10.1016/j.neuron.2012.01.010
- Eckstein, D., & Henson, R. N. (2012). Stimulus/response learning in masked congruency priming of faces: Evidence for covert mental classifications? *The Quarterly Journal of Experimental Psychology*, 65(1), 92-120. doi:10.1080/17470218.2011.590595
- Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When Categories Collide: Accumulation of Information About Multiple Categories in Rapid Scene Perception. *Psychological Science*, 22(6), 739-746. doi:10.1177/0956797611407930
- Fairhall, S. L., Anzellotti, S., Pajtas, P. E., & Caramazza, A. (2011). Concordance between perceptual and categorical repetition effects in the ventral visual stream. *Journal of Neurophysiology*, 106(1), 398-408. doi:10.1152/jn.01138.2010
- Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, 42(6), 790-796.
- Fitzgerald, J. K., Freedman, D. J., & Assad, J. A. (2011). Generalized associative representations in parietal cortex. *Nature Neuroscience*, 14(8), 1075-1079. doi:10.1038/nn.2878
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science (New York, N.Y.)*, 291(5502), 312-316. doi:10.1126/science.291.5502.312
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(12), 5235-5246.
- Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32, 209-224. doi:10.1146/annurev.neuro.051508.135603
- Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 36(4), 539-559.
- Glaser, W. R., & D\"ungelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology. Human Perception and Performance*, 10(5), 640-654.

- Hantsch, A., Jescheniak, J. D., & Mädebach, A. (2012). Naming and categorizing objects: Task differences modulate the polarity of semantic effects in the picture–word interference paradigm. *Memory & Cognition*, *40*(5), 760–768. doi:10.3758/s13421-012-0184-6
- Harris, A., & Nakayama, K. (2007). Rapid face-selective adaptation of an early extrastriate component in MEG. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(1), 63–70. doi:10.1093/cercor/bhj124
- Horner, A. J., & Henson, R. N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia*, *46*(7), 1979–1991. doi:10.1016/j.neuropsychologia.2008.01.018
- Horner, A. J., & Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 757–779. doi:10.1037/a0015262
- Horner, A. J., & Henson, R. N. (2011a). Incongruent Abstract Stimulus–Response Bindings Result in Response Interference: fMRI and EEG Evidence from Visual Object Classification Priming. *Journal of Cognitive Neuroscience*, *24*(3), 760–773. doi:10.1162/jocn_a_00163
- Horner, A. J., & Henson, R. N. (2011b). Stimulus-response bindings code both abstract and specific representations of stimuli: evidence from a classification priming design that reverses multiple levels of response representation. *Memory & Cognition*, *39*(8), 1457–1471. doi:10.3758/s13421-011-0118-8
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science (New York, N.Y.)*, *310*(5749), 863–866. doi:10.1126/science.1117593
- Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324. doi:10.1016/j.tins.2007.05.001
- Jerbi, K., Ossandón, T., Hamamé, C. M., Senova, S., Dalal, S. S., Jung, J., ... Lachaux, J.-P. (2009). Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Human Brain Mapping*, *30*(6), 1758–1771. doi:10.1002/hbm.20750
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. doi:10.1016/j.visres.2007.09.013
- Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of Vision*, *12*(13), 16–16. doi:10.1167/12.13.16
- Keysers, C., Xiao, D.-K., Foldiak, P., & Perrett, D. I. (2005). Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology*, *22*(3), 316–332. doi:10.1080/02643290442000103
- Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. doi:10.1152/jn.00024.2007
- Kiesel, A., Kunde, W., & Hoffmann, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology*, *3*(1-2), 307.

- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, *46*(11), 1762-1776. doi:10.1016/j.visres.2005.10.002
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36* *ECVP Abstract Supplement*.
- Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, *39*(2), 184-199.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126-1141. doi:10.1016/j.neuron.2008.10.043
- Kuipers, J.-R., & La Heij, W. (2012). Congruency effects in conceptualizing for speech. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2155-2168. doi:10.1080/17470218.2012.684693
- Kuipers, J.-R., La Heij, W., & Costa, A. (2006). A further look at semantic context effects in language production: The role of response congruency. *Language and Cognitive Processes*, *21*(7-8), 892-919. doi:10.1080/016909600824211
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596-9601.
- Lisman, J. E., & Jensen, O. (2013). The θ - γ neural code. *Neuron*, *77*(6), 1002-1016. doi:10.1016/j.neuron.2013.03.007
- Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, *62*(2), 281-290. doi:10.1016/j.neuron.2009.02.025
- Liu, Z., & Richmond, B. J. (2000). Response Differences in Monkey TE and Perirhinal Cortex: Stimulus Association Related to Reward Schedules. *Journal of Neurophysiology*, *83*(3), 1677-1692.
- Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513-536. doi:10.1080/13506280902937606
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485-495. doi:10.3758/BF03198265
- Lupker, S. J., & Katz, A. N. (1981). Input, decision, and response factors in picture-word interference. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(4), 269.
- Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The Time-Course of Visual Categorizations: You Spot the Animal Faster than the Bird. *PLoS ONE*, *4*(6), e5927. doi:10.1371/journal.pone.0005927
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word

interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 503-535. doi:10.1037/0278-7393.33.3.503

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3), 1407-1419. doi:10.1152/jn.90248.2008

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science (New York, N.Y.)*, 254(5036), 1377-1379.

Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys Quickly Learn and Generalize Visual Categories without Lateral Prefrontal Cortex. *Neuron*, 66(4), 501-507. doi:10.1016/j.neuron.2010.04.010

Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision*, 12(11), 19-19. doi:10.1167/12.11.19

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in Psychology*, 4, 128. doi:10.3389/fpsyg.2013.00128

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117(1), 12-22. doi:10.1016/j.bandl.2010.09.013

Pan, X., & Sakagami, M. (2012). Category representation and generalization in the prefrontal cortex. *European Journal of Neuroscience*, 35(7), 1083-1091. doi:10.1111/j.1460-9568.2011.07981.x

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

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, 47(3), 329-342.

Poncet, M., & Fabre-Thorpe, M. (2014). Stimulus duration and diversity do not reverse the advantage for superordinate-level representations: the animal is seen before the bird. *The European Journal of Neuroscience*. doi:10.1111/ejn.12513

Poncet, M., Reddy, L., & Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision*, 12(1), 15. doi:10.1167/12.1.15

Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., & Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 19(8), 1806-1819. doi:10.1093/cercor/bhn210

Praß, M., Grimsen, C., König, M., & Fahle, M. (2013). Ultra rapid object categorization: effects of level, animacy and context. *PLoS One*, 8(6), e68051. doi:10.1371/journal.pone.0068051

- Rolls, E. T., & Tovée, M. J. (1994). Processing Speed in the Cerebral Cortex and the Neurophysiology of Visual Masking. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 257(1348), 9-15. doi:10.1098/rspb.1994.0087
- Rosinski, R. R. (1977). Picture-Word Interference Is Semantically Based. *Child Development*, 48(2), 643-647. doi:10.2307/1128667
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5(7), 629-630. doi:10.1038/nn866
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Research*, 44(9), 877-894. doi:10.1016/j.visres.2003.11.014
- Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, 18(1), 16-25. doi:10.1016/j.tics.2013.10.010
- Roy, J. E., Buschman, T. J., & Miller, E. K. (2014). pFC Neurons Reflect Categorical Decisions about Ambiguous Stimuli. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_00568
- Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying Object Categories from Event-Related EEG: Toward Decoding of Conceptual Representations. *PLoS ONE*, 5(12), e14465. doi:10.1371/journal.pone.0014465
- Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. D., & Schacter, D. L. (2003). Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *NeuroImage*, 19(3), 613-626.
- Singer, J., & Kreiman, G. (2009). Toward unmasking the dynamics of visual perception. *Neuron*, 64(4), 446-447. doi:10.1016/j.neuron.2009.11.008
- Swaminathan, S. K., & Freedman, D. J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature Neuroscience*, 15(2), 315-320. doi:10.1038/nn.3016
- Swaminathan, S. K., Masse, N. Y., & Freedman, D. J. (2013). A comparison of lateral and medial intraparietal areas during a visual categorization task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(32), 13157-13170. doi:10.1523/JNEUROSCI.5723-12.2013
- Theeuwes, J., & Burg, E. V. der. (2013). Priming makes a stimulus more salient. *Journal of Vision*, 13(3). doi:10.1167/13.3.21
- Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, 23(3), 513-522.
- Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking Categories in the Brain. *Science*, 291(5502), 260-263. doi:10.1126/science.1058249
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520-522.

Tovée, M. J. (1994). Neuronal processing. How fast is the speed of thought? *Current Biology: CB*, 4(12), 1125-1127.

Ungerleider, L. G., Gaffan, D., & Pelak, V. S. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, 76(3), 473-484.

Van den Bussche, E., Notebaert, K., & Reynvoet, B. (2009). Masked Primes Can Be Genuinely Semantically Processed. *Experimental Psychology (formerly « Zeitschrift für Experimentelle Psychologie »)*, 56(5), 295-300. doi:10.1027/1618-3169.56.5.295

Vidal, J. R., Ossandón, T., Jerbi, K., Dalal, S. S., Minotti, L., Ryvlin, P., ... Lachaux, J.-P. (2010). Category-Specific Visual Responses: An Intracranial Study Comparing Gamma, Beta, Alpha, and ERP Response Selectivity. *Frontiers in Human Neuroscience*, 4, 195. doi:10.3389/fnhum.2010.00195

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex (New York, N.Y.: 1991)*, 4(5), 470-483.

Weiner, K. S., Sayres, R., Vinberg, J., & Grill-Spector, K. (2010). fMRI-Adaptation and Category Selectivity in Human Ventral Temporal Cortex: Regional Differences Across Time Scales. *Journal of Neurophysiology*, 103(6), 3349-3365. doi:10.1152/jn.01108.2009

5.3. Discussion

Les résultats comportementaux⁴ montrent un effet de congruence clair : si l'amorce et la cible appartiennent à la même catégorie cible (oiseau dans l'expérience 1 et animal dans l'expérience 2), alors les performances (temps de réponse et nombre de réponses correctes) sont meilleures que si l'amorce appartient à une catégorie différente de celle de la cible (l'amorce induit une réponse « non » alors que la cible induit une réponse « oui »). Ceci signifie que même si on demandait aux participants d'ignorer l'amorce, elle était automatiquement traitée et influençait le traitement de la cible. Cet effet de congruence n'est pas surprenant étant donné qu'il a déjà été montré dans des paradigmes de masquage où les participants ne sont même pas capables d'identifier l'amorce si on le leur demande (e.g. Dehaene et al., 2001). Cependant par rapport à ce type de paradigme, l'effet d'amorçage était beaucoup plus important dans notre étude puisque les temps de réponses médians étaient en moyenne 36 à 48 ms plus rapides dans les essais congruents que dans les essais incongruents.

Dans cette étude, nous considérons l'effet d'amorçage comme la différence entre la condition congruente et la condition incongruente. Malheureusement, il est impossible de déterminer si une amorce congruente a un effet facilitateur ou si une amorce incongruente a un effet d'interférence puisque nous n'avons pas de condition neutre, c'est-à-dire une amorce qui ne soit ni congruente ni incongruente. Il aurait pu être envisagé d'utiliser une amorce sans signification (texture) mais étant donné la tâche, oiseau/non-oiseau dans l'expérience 1 ou animal/non-animal dans l'expérience 2, même ce type d'amorce élicite une réponse « non ». Par ailleurs, ne pas utiliser d'amorce du tout revient au même type de tâche qui a été utilisé dans les autres études de catégorisation présentées dans cette thèse, en particulier dans l'étude 1 pour des réponses oui/non dans des tâches animal/non-animal et oiseau/non-oiseau. Si on compare les performances de cette étude avec celles obtenues ici, alors que ce soit en condition congruente ou en condition incongruente, les participants étaient plus lents et faisaient plus d'erreurs dans le paradigme d'amorçage que dans le protocole de catégorisation simple (sans amorce). La présentation d'une image distractive (amorce) juste avant une image cible diminue donc les performances en général peut être par des mécanismes d'inhibition divers (e.g. inhibition de

⁴ Dans l'article, seulement les résultats pour les réponses « oui » sont rapportées. Les résultats des réponses « non » sont analysés en annexe 8.3.

la prise de décision, inhibition des autres représentations de catégories) ou par une augmentation du « bruit » dans les patterns d'activité neuronale.

Par ailleurs, l'intérêt de cette étude repose sur l'utilisation d'amorces incongruentes plus ou moins sémantiquement proches de la cible. Nos résultats montrent que plus l'amorce et la cible ont des caractéristiques communes, plus l'effet d'amorçage (incongruent - congruent) est important, c'est-à-dire plus l'amorce interfère avec la réponse des participants. Ceci est vrai même quand l'amorce et la cible ne partagent pas la même catégorie superordonnée, c'est-à-dire, même si le lien sémantique entre l'amorce et la cible n'est pas pertinent pour la réalisation de la tâche (connaître la classe de l'objet, air/eau/sol, n'est pas utile pour catégoriser un animal).

Au contraire, dans les paradigmes d'interférence image-mot, lorsque les mots utilisés ont le même degré de convergence avec l'image, par exemple l'image d'un saxophone avec les mots trompette ou banjo, alors plus le mot est sémantiquement proche de l'image plus les performances sont élevées (Mahon et al., 2007). C'est-à-dire, le temps pour nommer l'image du saxophone est plus court si l'image est présentée avec le mot trompette qu'avec le mot banjo. Appliqué à l'expérience 2 de notre étude, c'est-à-dire quand les deux types d'amorces utilisées sont des véhicules, l'effet d'amorçage devrait être moins important pour des images sémantiquement proches (avion et oiseau) que pour des images sémantiquement éloignés (voiture et oiseau). Or, nous obtenons des résultats opposés : les participants étaient plus lents pour les amorces incongruentes partageant la même classe que la cible que pour les amorces incongruentes ne partageant pas la même classe que la cible (effet d'amorçage plus élevé dans le premier que dans le second cas).

Cet effet mérite d'être testé plus en détail afin de savoir s'il est la conséquence du partage (a) de caractéristiques visuelles, comme la forme globale de l'objet ou l'arrangement de certains traits visuels, (b) de caractéristiques conceptuelles ou fonctionnelles, comme la possibilité de voler, (c) d'un contexte similaire pour les deux objets (suivant la classe de l'objet, la probabilité de le voir apparaître dans un environnement particulier est plus ou moins élevée ; voir par exemple les résultats de Palmer, 1975). La première possibilité pourrait être testée en utilisant une technique de transformation graduelle entre un objet et un autre (« morphing »). Cependant, si on considère les résultats de l'expérience 2, il est difficile de trouver des exemples d'un partage de caractéristiques visuelles pour des objets de même classe vs. des objets de classes différentes et cette hypothèse est donc difficilement valide et exploitable.

La possibilité d'un effet du contexte peut être testée facilement. Par exemple, si on utilise des objets isolés, c'est-à-dire sans contexte, et que nous obtenons toujours un effet d'amorçage plus important pour une amorce avion que pour une amorce voiture sur une cible oiseau alors l'effet n'est pas dû au contexte de l'objet.

Dans les expériences d'amorçage utilisant des amorces traitées inconsciemment (car fortement masquées), il a parfois été argumenté que l'effet de congruence observé pourrait s'expliquer par une association dans la prise de décision entre un stimulus et une réponse (Denkinger & Koutstaal, 2009; D. Eckstein & Henson, 2012; Horner & Henson, 2011a, 2011b). Par exemple, dans notre expérience, la catégorie oiseau pourrait très rapidement activer la réponse « oui ». Cependant dans notre étude, les amorces n'étaient pas masquées mais surtout, nous observons une différence d'effet d'amorçage pour des amorces ayant la même association stimulus-réponse (réponse « non » pour les amorces non-oiseau que ce soient d'autres animaux ou des véhicules). Ainsi, nous interprétons nos résultats comme la conséquence d'une compétition entre l'activité évoquée par l'amorce et celle par la cible : plus les deux images sont sémantiquement proches et plus elles interfèrent l'une avec l'autre. Etant donné que cette compétition dépend de la relation sémantique entre l'amorce et la cible même pour des SOAs de 80 ms, nous proposons que cette compétition aurait lieu dans IT.

A partir de nos résultats mais aussi en considérant ceux obtenus en IRMf et en électrophysiologie, nous développons dans l'article 4 un modèle essayant d'expliquer les mécanismes de la catégorisation d'objets visuels. Si on considère que les caractéristiques visuelles sont codées par l'activation d'un réseau distribué de neurones dans IT (ce qui semble plausible d'après les résultats de Kriegeskorte et al., 2008 mais aussi d'après la théorie PDP de McClelland & Rogers, 2003), alors les patterns d'activité de deux objets se superposent plus lorsqu'ils partagent des caractéristiques que lorsqu'ils n'en partagent pas. L'idée développée dans le modèle est que, plus il y a de chevauchement, plus les deux patterns sont difficiles à différencier, ce qui rend la prise de décision plus difficile (plus longue et/ou induisant plus d'erreurs).

Nous suggérons dans l'article que la compétition entre différentes catégories se déroulerait dans IT. Cependant, nous ne sommes pas opposés au fait qu'elle puisse avoir lieu dans d'autres aires cérébrales, en particulier dans le cas de la deuxième expérience dans laquelle les caractéristiques partagées entre les deux images pourraient être d'ordre beaucoup plus conceptuel (e.g. objet qui peut voler) que visuel

(ressemblance physique en générale). Quoi qu'il en soit, cela ne change pas le message principal de cet article qui est que les activités évoquées par deux stimuli entrent en compétition et que le niveau de compétition dépend du niveau de similarité entre les stimuli.

Bien que l'idée que les caractéristiques visuelles pourraient être codées par un réseau de neurones ne soit pas nouvelle, aucune étude n'a auparavant testé l'effet d'un ensemble d'attributs actifs dans IT lors du traitement d'un autre objet. Je pense que ce paradigme peut s'avérer être un outil utile pour des recherches futures, en particulier parce qu'il peut être utilisé facilement pour déterminer la similarité entre des objets visuels sans utiliser des techniques telles que l'IRMf et l'électrophysiologie. Même si cela demande à être vérifié, il pourrait aussi être potentiellement utilisé pour tester des similarités plus conceptuelles entre les représentations.



Chapitre 6
Discussion



6. Discussion générale

L'objectif de cette thèse était de caractériser les processus de la catégorisation d'objets visuels afin de mieux comprendre comment ces mécanismes opèrent. De nombreuses études ont été réalisées sur ce sujet et de nouvelles sont publiées chaque jour. En effet, la capacité de catégoriser des objets est à la base de la reconnaissance visuelle puisqu'elle nous permet de reconnaître des objets sans que nous ayons à tous les mémoriser (et donc à surcharger notre mémoire) en les classant comme appartenant à telle ou telle catégorie et étant différent de telle ou telle autre.

J'ai d'abord présenté les publications qui m'ont le plus interpellées ou qui m'ont semblé les plus importantes dans le cadre de ma problématique. Cette revue n'est certainement pas exhaustive mais je pense qu'elle permet de donner un historique rapide du développement des hypothèses concernant la catégorisation d'objets et son implémentation neuronale. Il m'a semblé important de faire cet historique afin de comprendre non seulement le contexte actuel dans lequel s'inscrivent mes travaux, mais aussi de savoir quel a été le développement des idées sur ce sujet.

J'aimerais signaler que les stimuli utilisés dans chacune des études de cette thèse étaient des scènes naturelles complexes (photographies de notre environnement) qui n'étaient vues qu'une seule fois par les participants. En effet, afin de comprendre les mécanismes cognitifs opérant dans notre vie quotidienne, il m'a semblé important de n'utiliser que des scènes naturelles au contraire d'objets visuels isolés ou de dessins (stimuli souvent utilisés et répétés plusieurs fois dans d'autres expériences). L'intérêt de la reconnaissance visuelle est aussi et surtout d'être capable de comprendre, en un temps imperceptible pour l'humain, un environnement naturel complexe. J'ai pu montrer dans les résultats de quatre études que la catégorisation d'objet était très efficace (rapide et automatique) dans le cas des objets de notre environnement.

Afin de rendre ce chapitre de discussion générale le plus clair possible, je résume tout d'abord les résultats des quatre articles de cette thèse dans le chapitre 6.1 et je développe ensuite leurs implications de façon plus générale dans un deuxième chapitre 6.2. Les deux chapitres suivant discutent des limitations de mes résultats ainsi que des perspectives futures pouvant être envisagées afin de mieux comprendre l'organisation et le fonctionnement des mécanismes de la catégorisation d'objets.

6.1. Résumé des résultats

6.1.1. Etude 1

Dans le premier article de cette thèse, j'ai voulu comprendre la différence du traitement d'un objet selon qu'il était catégorisé à différents niveaux d'abstraction. En effet, pour une même image de canari, il est possible de le reconnaître en tant que animal (catégorie superordonnée, abstraite) ou en tant que oiseau (catégorie basique, moins abstraite) ou en tant que passereau (ou oiseau chanteur, catégorie subordonnée). L'accès à différents niveaux conceptuels pourrait faire intervenir différents traitements visuels. En fait, depuis Rosch et al., (1976), il était communément accepté que l'accès aux catégories d'objets se faisait d'abord au niveau basique avant d'être réalisé au niveau superordonné et encore plus tard au niveau subordonné. Cependant, de récents articles remettaient cette idée en question (e.g. Macé et al., 2009; McClelland & Rogers, 2003).

Dans l'article 1, j'ai ainsi testé la vitesse de catégorisation de divers objets visuels pour différents niveaux d'abstraction tout en faisant varier certains paramètres tels que le temps de présentation des images, la diversité des stimuli au sein des catégories, ou le mode de réponse. Un des résultats clairs de cet article est que le temps de présentation (20, 250, ou 500 ms) n'a aucun effet sur les performances de catégorisation (expérience 1). Un deuxième résultat clair est que les performances sont meilleures en go/no-go que dans une tâche oui/non mais la diminution des performances (temps de réponse et précision) dans cette seconde tâche est la même quel que soit le niveau d'abstraction requis par la tâche (expérience 4).

Il semblerait que les deux paramètres (a) niveau de catégorisation accédé le plus rapidement, et (b) diversité des stimuli au sein des catégories, soient liés. En effet, même si nos résultats montrent que la catégorisation animal/non-animal est mieux réussie et est réalisée avant la catégorisation oiseau/animal non-oiseau, les résultats sont moins clairs pour la catégorisation de véhicules. Par exemple, la catégorisation moto/véhicule non-moto est aussi rapide et efficace que la catégorisation véhicule/non-véhicule. Cependant, la catégorisation voiture/véhicule non-voiture est plus lente (bien que le d' soit le même) que la catégorisation véhicule/non-véhicule. Par ailleurs, plus les exemplaires au sein des catégories cibles et distracteurs sont variés, plus les performances de catégorisation diminuent. Ainsi dans l'article, nous émettons l'hypothèse que les bonnes performances en catégorisation

moto/véhicule non-moto sont dues à l'homogénéité de la catégorie moto et à sa différence claire par rapport aux autres véhicules. Les performances de catégorisation seraient alors déterminées par les caractéristiques diagnostiques nécessaires à la réalisation de la tâche.

Dans l'article nous émettons deux possibilités pour expliquer ces résultats. La première serait que les aires cérébrales engagées dans les processus de décision changeraient leur « lecture » des aires visuelles en fonction des caractéristiques diagnostiques dictées par la tâche (traits visuels permettant de distinguer la catégorie cible de la catégorie distracteur) (e.g. S. Li, Mayhew, & Kourtzi, 2009). Une autre hypothèse est que la sélectivité des populations de neurones des aires visuelles pourrait être ajustée en fonction de la tâche par des mécanismes top-down probablement provenant des aires frontales (Miller, Vytlačil, Fegen, Pradhan, & D'Esposito, 2011; Reynolds & Heeger, 2009). Une hypothèse n'exclut pas l'autre et il serait possible que ces deux mécanismes aient tous deux un rôle dans la catégorisation visuelle.

Globalement, ces résultats suggèrent qu'il ne faudrait pas considérer les différents niveaux d'abstraction comme des traitements différents ou une hiérarchie entre les différents niveaux de catégorisation mais la conséquence d'une compétition entre cibles et distracteurs (voir aussi les résultats de l'article 4).

6.1.2. Etude 2

Dans la deuxième étude, nous avons voulu tester les besoins en attention pour les catégorisations basiques. L'article de F. F. Li et al., (2002) a montré qu'il était possible de réaliser des catégorisations superordonnées telles que animal/non-animal et véhicule/non-véhicule avec des ressources attentionnelles minimales. Dans l'article 1, nous avons montré que les catégorisations basiques oiseau/animal non-oiseau et voiture/véhicule non-voiture étaient réalisées moins efficacement (temps de réponse en particulier) que les catégorisations superordonnées animal/non-animal et véhicule/non-véhicule. Ainsi, les catégorisations basiques nécessitent un niveau de traitement plus en détail que les catégorisations superordonnées et ce niveau de traitement pourrait requérir de l'attention.

Pour tester cette hypothèse, nous avons utilisé le même paradigme de double-tâche que F. F. Li et al., (2002). Nos résultats montrent clairement que les performances en catégorisation superordonnée et basique sont aussi bonnes lorsque la tâche de catégorisation est réalisée seule, c'est-à-dire lorsque l'attention est disponible, que lorsqu'elle est réalisée simultanément avec une tâche centrale qui requiert la majorité des ressources attentionnelles. Ainsi, même s'il est plus long de catégoriser un objet en tant qu'oiseau qu'en tant qu'animal, ces deux niveaux d'abstraction ont les mêmes besoins en attention, c'est-à-dire des besoins quasi-inexistants.

Pourtant, nous observons bien une différence entre les tâches de catégorisation au niveau superordonné et au niveau basique. En effet, pour le même pourcentage de réponses correctes, la scène naturelle devait être présentée pendant 85 ms et 87 ms pour la catégorisation animal et véhicule respectivement, alors qu'elle devait être présentée pendant 180 ms et 140 ms pour les catégorisations chien et voiture (il faut noter que dans ce paradigme les images étaient masquées, au contraire du paradigme utilisé dans l'article 1). En fait, si on utilise des temps de présentation plus courts, alors le pourcentage de réponses correctes diminue même dans la condition où la tâche est réalisée seule, c'est-à-dire lorsque toutes les ressources attentionnelles peuvent y être allouées.

Ces résultats montrent que la réalisation d'une tâche au niveau basique demande plus temps de présentation, c'est-à-dire plus de temps pour accéder à certaines informations que la réalisation d'une tâche au niveau superordonné.

6.1.3. Etude 3

La catégorisation d'objets visuels semble être réalisée sans efforts et de fait, elle peut être réalisée avec un minimum d'attention (article 2). La question qui se pose lorsqu'on réalise ce type de tâche est jusqu'à quel niveau de détail les images sont-elles traitées?

Dans l'article 3, nous avons testé si des images vues dans une tâche de catégorisation animal/non-animal peuvent être reconnues ultérieurement en demandant à des participants de faire une tâche familier/non-familier. Le set de 200 images familières pouvaient soit avoir été vu une seule fois avant la

tâche de reconnaissance (groupe non-entraîné), soit avoir été vu une fois par jour pendant trois semaines (groupe entraîné).

Dans la tâche de reconnaissance, les 200 images familières (vues 21 fois par le groupe entraîné et seulement une fois par le groupe non-entraîné) étaient présentées 20 ms au milieu d'un set de 200 images du même type (scènes naturelles provenant de la même base de données). Les 200 images familières étaient répétées trois fois avec 200 images nouvelles à chaque répétition (donc au total, 600 images nouvelles ont été utilisées). Les participants étaient testés dans un paradigme de go/no-go : ils devaient lever le doigt le plus rapidement s'ils avaient un sentiment de familiarité avec les images, et leur EEG était enregistré.

Les participants entraînés étaient capables dès la première répétition de réaliser la tâche familier/non-familier en 550 ms avec un d' de 2.5. Ces résultats indiquent que durant leur entraînement sur la tâche de catégorisation animal/non-animal, les images présentées seulement 20 ms ont été mémorisées même si les participants ne savaient pas qu'ils allaient être testés plus tard dans une tâche de reconnaissance. Pourtant une trace mnésique implicite a bien été construite au cours de la répétition des stimuli. La différence entre les images familières et nouvelles était aussi visible dans les tracés EEG, avec une différence entre les ERPs de ces deux sets d'images débutant vers 230 ms après l'apparition du stimulus.

Les participants non-entraînés, quant à eux, étaient à peine capables de réaliser la tâche familier/non-familier lors de la première répétition des images ($d'=0.5$) mais leurs performances augmentaient avec le nombre de répétitions (suivant les participants, on observe une courbe d'apprentissage très forte ou pas d'apprentissage du tout). Les résultats en EEG ne montrent aucune différence entre les potentiels évoqués pour les images familières et ceux pour les images nouvelles.

Un autre résultat intéressant qu'apporte cette étude est que, même si les participants réalisaient une tâche de familiarité, une différence entre les catégories animal et non-animal était observée en ERP. De par sa latence (début vers 150ms) et sa localisation (électrodes frontales principalement), cette différence d'activité correspond assez bien à celle observée dans d'autres études où les participants devaient réaliser explicitement une tâche de catégorisation animal/non-animal. Ainsi, l'information sur la catégorie de l'objet, ou en tout cas si l'objet est un animal ou pas, serait traitée automatiquement.

6.1.4. Etude 4

Une des questions qui n'a jamais été vraiment posée dans le domaine de la catégorisation visuelle est celle de l'interaction entre différentes catégories d'objets. C'est cette problématique qui est développée dans l'article 4 en testant l'influence d'un ensemble de caractéristiques visuelles actives dans IT lors du traitement d'une nouvelle image.

Pour tester les interactions entre différentes catégories d'objets nous avons utilisé un paradigme d'amorçage : une amorce flashée seulement 20 ms et devant être ignorée était présentée 80 ou 180 ms avant une image cible présentée pendant 100 ms. Les participants devaient catégoriser l'image cible au niveau basique (expérience 1) ou au niveau superordonné (expérience 2). La particularité de cette étude était de tester différents types d'amorces : elles pouvaient être soit sémantiquement proches de la cible (e.g. un chien et un oiseau), soit sémantiquement éloignées (e.g. une voiture et un oiseau).

Nos résultats suggèrent que les activités évoquées par l'amorce et la cible entrent en compétition. Cette compétition est vite résolue (ou il n'y a peut-être même pas de compétition) si l'amorce et la cible induisent la même décision (et réponse motrice). Cependant, dans le cas où la cible et l'amorce induisent des réponses contraires (condition incongruente), alors plus les deux objets sont sémantiquement proches (et donc partagent des caractéristiques telle que « être un animal »), plus la compétition est importante.

Après la présentation de la cible, le pattern d'activité de IT représenterait à la fois l'amorce et la cible. Selon différentes études, et notamment celle de Kriegeskorte et al., (2008), les patterns d'activité de IT seraient plus similaires pour des objets similaires que pour des objets plus éloignés. Nous proposons dans notre modèle que les deux patterns d'activité évoqués par l'amorce et la cible seraient plus difficilement dissociés lorsque les deux objets partagent des caractéristiques que lorsqu'ils n'en partagent pas. Il y a néanmoins d'autres hypothèses envisageables, par exemple :

1. Etant donné que l'amorce doit être ignorée, on pourrait imaginer un mécanisme d'inhibition de l'activité provoquée par la présentation de celle-ci. Dans ce cas, les populations neuronales codant pour des caractéristiques communes entre les deux images mettraient plus de temps à être réactivées lors de la présentation de la cible. Ce mécanisme pourrait expliquer un effet d'amorçage plus important pour les amorces proches que pour les amorces éloignées de la

cible. Cependant, cette hypothèse se heurte à un problème majeur : les temps de réponses devraient être les plus lents dans le cas d'un essai congruent que dans celui d'un essai incongruent puisque la majorité des caractéristiques sont communes aux deux images dans la condition congruente. Au contraire, c'est dans cette condition que les temps de réaction sont les plus rapides.

2. Pour pallier le problème de cette première hypothèse, on pourrait imaginer l'existence d'un mécanisme qui vérifierait s'il y a un conflit au niveau de la prise de décision après la présentation de la cible. S'il n'y a pas de conflit (essai congruent), alors le traitement visuel suivrait une voie particulière permettant de réaliser la tâche très rapidement. Dans le cas où il y aurait un conflit (essai incongruent), alors une autre voie de traitement entrerait en jeu. Etant donné qu'on observe une différence entre différents types d'amorces incongruentes, alors cette étape supplémentaire doit faire intervenir un mécanisme dépendant de la similarité entre les patterns de réponses, ce qui revient à l'hypothèse développée dans notre modèle.
3. Un autre phénomène d'inhibition pourrait avoir lieu, pas dans le sens où l'amorce doit être ignorée, mais par une inhibition latérale des concepts sémantiquement proches lors de l'apparition d'un objet (ici de l'amorce). Par exemple, dans le cas d'une amorce chien dans l'expérience 1, le système inhiberait les représentations des autres catégories d'animaux, dont les oiseaux, tandis que dans le cas d'une amorce véhicule, l'inhibition ne s'étendrait pas ou très peu aux oiseaux. Ainsi, l'effet d'amorçage pour les amorces éloignées (véhicule) pourrait s'expliquer par un conflit au niveau de l'association stimulus-réponse alors que celui pour les amorces proches (animal non-oiseau) s'expliquerait par le même conflit stimulus-réponse auquel s'ajouterait un délai correspondant à la réactivation de la catégorie oiseau inhibée par l'amorce. L'effet de cette inhibition latérale serait donc beaucoup moins important que l'effet du conflit stimulus-réponse, mais pourrait augmenter avec de plus longs SOAs. Ceci expliquerait que dans nos résultats, l'effet d'amorçage augmente avec l'augmentation du SOA mais seulement dans le cas des amorces proches (le conflit stimulus-réponse reste le même). Dans la deuxième expérience (catégorisation animal/non-animal), du fait de l'association stimulus-réponse les performances sont meilleures dans la condition congruente que dans la condition incongruente (résultat d'un conflit stimulus-réponse). Avec l'augmentation du SOA,

l'inhibition latérale pourrait s'amplifier et atteindre des catégories conceptuellement proches telles que « peut voler ». Ainsi, à un SOA court, une amorce conceptuellement proche n'inhiberait pas une cible oiseau mais ce serait le cas à un SOA plus long. C'est-à-dire qu'en complément d'un conflit stimulus-réponse, il y aurait aussi une inhibition de la cible oiseau pour des SOAs longs mais seulement dans le cas d'une amorce conceptuellement proche.

L'hypothèse développée dans notre modèle explique bien nos résultats, au contraire de la première hypothèse. Elle est aussi plus simple que la deuxième hypothèse qui fait intervenir un mécanisme additionnel. Concernant la troisième hypothèse, même si elle est plausible, c'est aussi la plus incertaine. En effet, même si des phénomènes d'inhibition latérale ont été montrés à un niveau très local (Mruczek & Sheinberg, 2012; Wang, Fujita, & Murayama, 2000), c'est-à-dire entre quelques neurones, il n'y a aucune raison de croire à un mécanisme d'inhibition latérale entre des concepts tels que « animal » ou « véhicule » qui émergeraient de l'activité combinée de populations de neurones. En fait, il est au contraire souvent observé, en particulier dans les études sur le langage, qu'un concept active d'autres concepts similaires. Ainsi, l'hypothèse développée dans l'article 4 semble être la plus appropriée.

Les résultats de cet article supportent l'idée que la catégorisation d'objets serait basée sur la possession (et/ou la non-possession) de certaines caractéristiques diagnostiques, qui seraient représentées sur un espace sémantique continu. C'est principalement cette hypothèse qui sera utilisée pour développer le prochain chapitre sur l'implication des résultats de cette thèse.

6.2. Implications

Les données collectées au cours de cette thèse soutiennent l'hypothèse que les objets seraient représentés dans IT en fonction de leurs caractéristiques. Ces caractéristiques seraient codées de façon automatique (article 3) et avec un minimum d'attention (article 2) aboutissant à une représentation distribuée et un espace continu des catégories d'objets dans IT. Une fois représentées dans IT, ces caractéristiques pourraient être « lues » afin de réaliser des tâches de catégorisation.

Ce mécanisme de lecture permettrait de discriminer l'activité des populations de neurones de IT provoquée par un stimulus et le « bruit » ambiant (activité aléatoire des populations de neurones), un peu dans l'idée de la théorie de détection de signal. Si plusieurs caractéristiques sont actives simultanément mais ne correspondent pas à la même catégorie d'objet (comme c'est le cas dans l'article 4 pour les essais incongruents) alors une compétition entre l'amorce et la cible a lieu. Dans le cas de la catégorisation d'un objet présenté seul (article 1), alors la compétition s'opère entre l'activité attendue (dépendant de la catégorie cible de la tâche) et les distracteurs. Que deux images ou qu'une seule image soit présentée, les performances de catégorisation vont dépendre de la similarité entre les caractéristiques de la cible et des distracteurs (cette hypothèse explique aussi les différences de performances suivant le niveau de catégorisation requis pour faire la tâche).

Dans l'article 1, j'ai pu montrer que les performances de catégorisation dépendaient de la similarité entre les exemplaires de la catégorie cible et de la catégorie distracteur (expérience 3 de l'article 1 en particulier). D'après le modèle développé dans l'article 4, ces résultats pourraient s'expliquer par :

1. les caractéristiques requises pour catégoriser l'objet au niveau demandé, c'est-à-dire le pattern d'activité attendu pour la catégorie cible
2. le chevauchement de certaines caractéristiques avec celles de certains distracteurs, en d'autres termes, le chevauchement du pattern d'activité des distracteurs et celui du pattern d'activité attendu pour la catégorie cible.

Ainsi, les temps de réponses sont plus rapides pour une catégorie cible incluant des exemplaires très similaires puisque les caractéristiques (le pattern d'activité) attendues sont bien définies et moindres. Le nombre d'erreur augmente avec l'augmentation du nombre de distracteurs puisque la probabilité

qu'un distracteur partage des caractéristiques communes (un pattern d'activité similaire) à celles de la catégorie cible augmente.

Ainsi, il ne faudrait pas considérer les différents niveaux de catégorisation (superordonné, basique, subordonné) comme des modules différents ou des étapes différentes de traitement, mais comme dépendants des caractéristiques devant être vérifiées avant de pouvoir catégoriser un objet à tel ou tel niveau d'abstraction. Par exemple, catégoriser un oiseau demande de vérifier plus de caractéristiques distinctives d'autres animaux (ou au moins pas les mêmes) que de catégoriser un oiseau par rapport à des véhicules et ainsi les temps de réponses sont plus longs dans le premier que dans le second cas. Il semble donc beaucoup plus approprié de parler en termes de caractéristiques qu'en termes de niveau d'entrée (voir chapitre 2.1). En fait, l'ensemble de la voie visuelle ventrale montre un effet d'adaptation à la catégorie basique ou superordonnée d'un objet (Chao, Weisberg, & Martin, 2002; Fairhall et al., 2011; Koutstaal et al., 2001; Pourtois et al., 2009; Simons et al., 2003; Weiner et al., 2010). Il ne semble donc pas y avoir différentes régions dédiées au traitement de différents niveaux de catégorisation.

Les caractéristiques requises pour réaliser une tâche au niveau basique (oiseau/animal non-oiseau) entraînent un décalage temporel par rapport aux caractéristiques requises pour réaliser une tâche au niveau superordonné (animal/non-animal). De plus, d'après les résultats de l'article 2, un temps de présentation plus long est nécessaire pour catégoriser des images au niveau basique qu'au niveau superordonné. Il semblerait donc que la lecture des caractéristiques activées lors de la présentation de l'image ait besoin d'être réalisée avant l'arrivée de l'activité évoquée par le masque. Ce décalage dans les temps de réaction et du temps de présentation pourrait être expliqué soit par (1) l'activation de caractéristiques nécessaires à la catégorisation basique plus tardive que l'activation de caractéristiques nécessaires à la catégorisation superordonnée, ou par (2) un nombre plus élevé de caractéristiques à vérifier au niveau basique. Ces deux possibilités ainsi qu'une troisième sont développées dans le chapitre suivant (6.2.1). D'autres hypothèses visant à expliquer comment les caractéristiques pourraient être codées dans IT et un possible rôle de l'attention sont développées dans les chapitres 6.2.2 et 6.2.3.

6.2.1. D'où vient le décalage temporel ?

La différence du temps de réponse suivant les caractéristiques nécessaires dans la réalisation d'une tâche de catégorisation pourrait s'expliquer par trois raisons : (1) les caractéristiques sont activées à des latences différentes, les caractéristiques globales ou superordonnées étant activées avant les caractéristiques plus détaillées, (2) les caractéristiques sont activées toutes en même temps mais c'est la lecture de IT qui est différente suivant le nombre de caractéristiques devant être vérifiées, (3) les caractéristiques visuelles (e.g. a un bec) sont codées en même temps mais avant les caractéristiques plus conceptuelles (e.g. peut voler). Je développe ces trois hypothèses ci-dessous en les discutant d'après les résultats de cette thèse.

6.2.1.1 Caractéristiques codées séquentiellement

Toutes les caractéristiques pourraient être encodées ou activées de façon séquentielle dans la voie visuelle ventrale, IT en particulier. Ainsi, les caractéristiques globales d'un objet pourraient être activées avant ses caractéristiques plus précises. Cette hypothèse implique que les patterns d'activité dans IT soient dynamiques, c'est-à-dire qu'ils évoluent en fonction du temps. Cette idée semble tout à fait probable puisque après un certain temps, l'activité évoquée par un stimulus disparaît. Cette activation séquentielle des caractéristiques visuelles se manifeste par des temps de réaction plus rapides pour les catégorisations au niveau superordonné qu'aux niveaux basique ou subordonné dans l'article 1. Elle explique aussi pourquoi dans l'article 2 les catégorisations basiques demandent un temps de présentation des images masquées plus long que pour les catégorisations superordonnées.

C'est aussi cette hypothèse qui est proposée dans l'article 4. Au fur et à mesure du temps imparti entre la présentation de l'amorce et de la cible (SOA), de plus en plus de caractéristiques pourraient être encodées (et/ou plus fortement activées, et/ou avec un recrutement plus important de neurones). Ainsi dans l'expérience 1, l'effet d'amorçage augmente avec l'augmentation du SOA pour des amorces qui induisent une réponse différente mais qui sont cependant très proches de la cible. Cette hypothèse explique aussi très bien les résultats de l'expérience 2 dans laquelle on observe un effet de la distance entre l'amorce et la cible seulement à des SOAs longs, lorsque des caractéristiques plus conceptuelles (e.g. peut voler) seraient activées.

Cependant, si les caractéristiques basiques sont codées après les caractéristiques superordonnées, alors on devrait observer un effet d'amorçage apparaissant plus tardivement pour des catégorisations basiques que pour des catégorisations superordonnées. Au contraire, nos résultats montrent que pour le SOA le plus court, une amorce « chat » influence les performances de catégorisation sur une cible « oiseau » autant qu'une amorce « véhicule » sur une cible « animal ». En d'autres termes, un nombre suffisant de caractéristiques basiques de l'amorce ont été encodé pour interférer avec la cible. Toutefois, il est possible que la différence entre le temps pour encoder les caractéristiques basiques et celui pour encoder les caractéristiques superordonnées soit assez faible. Le paradigme utilisé dans l'article 4 pourrait ne pas être assez sensible pour faire cette différence.

Un des problèmes de cette hypothèse est qu'elle est difficile à concilier avec les résultats de classification obtenus pour différents types d'enregistrements de l'activité cérébrale. Par exemple, Hung et al. (2005) ont montré en enregistrant des populations des neurones de IT que toutes les informations nécessaires pour catégoriser (caractéristiques globales) ou identifier (caractéristiques propres à un stimulus donné) un stimulus étaient présentes dans la même fenêtre temporelle. D'autres résultats en EEG et MEG montrent eux que les informations permettant l'identification des objets seraient présentes avant celles permettant leur catégorisation (Carlson et al., 2013; Cichy et al., 2014).

6.2.1.2 Caractéristiques codées en même temps

Une autre hypothèse qui pourrait être avancée est que toutes les caractéristiques d'un objet seraient activées en même temps ou à des temps très similaires et c'est le temps de lecture de IT qui serait modifié suivant la tâche de catégorisation à réaliser. Ainsi, suivant le niveau de catégorisation et la similarité des exemplaires au sein des catégories d'objets utilisées, le nombre de caractéristiques devant être vérifiées (ou écartées) serait différent et entraînerait des temps de lecture de IT différents. En d'autres termes, toutes les caractéristiques sont activées rapidement mais la lecture des caractéristiques et la prise de décision qui s'ensuit sont retardées dans le cas d'une catégorisation mettant en jeu des objets similaires (e.g. catégorisation basique) par rapport à des objets plus faciles à différencier (e.g. catégorisation superordonnée).

Les temps de présentation nécessaires à différents types de catégorisation (article 2) peuvent aussi être expliqués par cette hypothèse. Le masque va entraîner l'activation et l'inhibition de populations de IT de façon aléatoire et il faut que le mécanisme de lecture ait le temps de lire les caractéristiques nécessaires à la catégorisation avant que le masque brouille le message contenu dans IT. De ce fait, les temps de présentation ont besoin d'être plus longs pour les catégorisations basiques que pour les catégorisations superordonnées.

Concernant maintenant les résultats de l'article 4, si on considère que toutes les caractéristiques d'un objet sont représentées au même moment dans IT alors on peut s'attendre à observer un effet de compétition entre une amorce et une cible quel que soit le niveau de catégorisation, même pour un SOA de 80 ms. Cependant, on observe un effet d'amorçage qui augmente dans l'expérience 1 et qui apparaît dans l'expérience 2 dans les essais où l'amorce et la cible sont incongruentes et proches. Comment expliquer ces résultats ?

Une possibilité serait que les caractéristiques *visuelles* seraient activées en même temps dans IT mais d'autres caractéristiques plus *conceptuelles* seraient activées plus tard dans d'autres aires cérébrales (e.g. cortex medio-parietal, cortex frontal). En effet, étant donné que les voxels du cortex occipital et inféro-temporal ont des organisations similaires pour les humains et les singes (Kriegeskorte et al., 2008), elles représentent probablement des caractéristiques visuelles et non pas des caractéristiques conceptuelles. Ainsi, « a un bec » serait codé dans IT alors que « peut voler » serait codé plus tardivement ailleurs que dans IT.

On pourrait alors imaginer deux lieux de compétition lorsque l'amorce et la cible partagent des caractéristiques mais pas la même réponse. Une première compétition aurait lieu au niveau du mécanisme de lecture de IT lorsque des caractéristiques visuelles sont partagées et une deuxième compétition, engageant d'autres aires cérébrales servant à coder des représentations plus conceptuelles et se produisant seulement pour des SOAs assez longs, pourrait s'ajouter lorsque des caractéristiques conceptuelles seraient partagées.

Etant donné que les temps de réponse des participants ne sont pas vraiment différents dans la condition avec un SOA de 80 ms et celle avec un SOA de 180 ms, il semble improbable qu'un mécanisme supplémentaire soit mis en jeu dans un cas par rapport à l'autre. Cependant, il est tout à

fait possible que le traitement des stimuli fasse intervenir ces deux niveaux de lecture pour les deux SOAs. Dans le cas d'un SOA court, il n'y aurait pas de conflit au niveau conceptuel, puisque l'activité causée par l'amorce n'aurait pas le temps de se propager à des régions représentant des caractéristiques conceptuelles. Dans le cas d'un SOA plus long, en plus d'une possible compétition au niveau de IT, cette activité aurait le temps d'atteindre les aires traitant les caractéristiques plus conceptuelles des objets.

6.2.1.3 *Caractéristiques codées en feed-forward et d'autres en feed-back*

On a considéré les caractéristiques comme étant codées lors d'une première vague automatique feed-forward. Pourtant, dans le cas d'un SOA de 180 ms, il est possible que des mécanismes de feed-back puissent influencer les patterns d'activité de IT avant le commencement du traitement de la cible dans IT. On pourrait alors imaginer que la compétition entre les caractéristiques des objets se déroule toujours au niveau de IT mais que les patterns d'activité soient modifiés par des feed-back. Ainsi, les caractéristiques *visuelles* pourraient être codées très rapidement et automatiquement lors d'une première vague feed-forward mais les caractéristiques *conceptuelles* seraient, elles, représentées dans d'autres aires cérébrales et modifieraient les patterns d'activité de IT par feed-back. Ces feed-back ne pourraient modifier l'activité des patterns de IT que si le traitement de la cible dans IT n'a pas déjà commencé.

Cette hypothèse est très proche de la précédente puisqu'elle suppose une activation simultanée des caractéristiques visuelles dans IT. La différence entre les temps de réponse dans une tâche de catégorisation serait causée par le temps mis par le mécanisme de lecture de IT pour être capable de distinguer le pattern d'activité de la cible par rapport au bruit (très élevé du fait de l'amorce). La différence avec l'hypothèse précédente est que dans celle-ci, il n'y a qu'un seul lieu de compétition et un seul mécanisme de lecture. Cependant, comme déjà mentionné dans l'hypothèse précédente, il semblerait que IT représenterait plutôt des caractéristiques visuelles et les patterns d'activité seraient difficilement distinguables pour des types de catégorisation autres que animé/non-animé et corps/visage. La nature des représentations dans IT est discutée mais pourrait être un mélange des propriétés physiques et conceptuelles des objets (Baldassi et al., 2013; Bell, Hadj-Bouziane, Frihauf, Tootell, & Ungerleider, 2008; Carlson et al., 2014; Harel, Kravitz, & Baker, 2014). Il est aussi possible

que nous n'ayons pas aujourd'hui les moyens techniques de faire la différence entre des patterns d'activité d'objets similaires (comme il était impossible avant les techniques d'analyses en MVPA de faire la différence entre des différences d'activité cérébrale pour différents objets).

6.2.1.4 Conclusion

Les résultats de cette thèse supportent l'idée que les performances de catégorisations dépendent des caractéristiques devant être vérifiées pour réaliser la tâche. Suivant ces caractéristiques diagnostiques, j'ai pu observer des temps de traitement, des temps de présentation et des effets d'amorçages différents. J'ai développé dans ce chapitre trois hypothèses pouvant expliquer ces résultats. D'autres expériences seront nécessaires afin de confirmer, infirmer ou modifier ces hypothèses.

6.2.1. Comment sont codées les caractéristiques ?

Ces hypothèses se basent sur l'idée que les caractéristiques sont codées de façon distribuée dans IT sur un espace sémantique continu (Çukur, Nishimoto, Huth, & Gallant, 2013; Huth et al., 2012). Par ailleurs, nos résultats montrent que ces caractéristiques, qui ne sont pas des traits visuels simples mais probablement un arrangement de traits complexes (voir aussi chapitre 6.6.3) sont traitées de façon automatique et en quasi-absence d'attention (résultats des études présentées dans les chapitres 3 et 4).

Ce traitement rapide, automatique et ne nécessitant que pas ou très peu de ressources attentionnelles pourrait s'expliquer par une modification de l'activité de la voie visuelle ventrale par des mécanismes de top-down mis en place en fonction de la tâche. En effet, plusieurs articles ont montré que les aires de la voie visuelle ventrale pouvaient être modulées en fonction des attentes des participants, que ce soit avant la présentation du stimulus ou lors de la présentation du stimulus (Chiu, Esterman, Han, Rosen, & Yantis, 2011; Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005; Harel et al., 2014; Kastner & Pinsky, 2004; Puri, Wojciulik, & Ranganath, 2009; Ranganath, DeGutis, & D'Esposito, 2004). Une autre possibilité est que la « lecture » des aires visuelles ne soit focalisée que sur des ensembles de traits particuliers qui seraient distinctifs de la catégorie cible par rapport à la catégorie distracteur. Ainsi, quel que soit le niveau de catégorisation, les participants seraient capables de réaliser la tâche

sans attention car peu de caractéristiques nécessiteraient d'être « lues » et le temps de présentation minimum (voir article 2) dépendrait du type de caractéristiques nécessitant d'être lu.

Cependant, ces deux hypothèses vont à l'encontre des résultats de l'étude 3. En effet dans cette expérience, les participants réalisaient une tâche familier/non-familier et non pas une tâche de catégorisation animal/non-animal. Pourtant, les caractéristiques possédées par l'objet étaient quand même traitées (différence d'activité en EEG pour les images d'animal vs. non-animal). De même, les résultats de Carlson et al. (2013) montrent qu'il est possible de déterminer la catégorie de l'objet vu par un participant d'après l'analyse des enregistrements en MEG même si le participant est en train de réaliser une tâche de discrimination voyelle/consonne. Un dernier résultat toujours en défaveur de ces deux hypothèses est que la tâche n'a que très peu d'effet sur les performances de classification de IT en fMRI (Carlson et al., 2003).

Les caractéristiques semblent donc être activées de façon automatique et en quasi-absence d'attention. Dans les premières étapes du traitement visuel, les neurones répondent à des traits visuels simples et il est nécessaire que ces traits soient intégrés en un même objet pour pouvoir le reconnaître. VanRullen (2009) a proposé que deux mécanismes d'intégration existeraient : un mécanisme feed-forward automatique reposant sur une connectivité neuronale spécifique aux objets de notre environnement, et un mécanisme sur-demande impliquant des processus de feed-back et qui, au contraire du précédent, nécessiterait des ressources attentionnelles (voir chapitre 2.1.3).

Cette connectivité neuronale permettant l'intégration automatique des traits visuels ainsi que l'organisation de l'ensemble de la voie visuelle ventrale pourraient être le résultat de l'expérience (Op de Beeck & Baker, 2010). Par exemple, la cooccurrence de certaines caractéristiques pourrait être responsable de la mise en place d'une intégration automatique des configurations familières de traits visuels (apprentissage non-supervisé e.g. McClelland & Rogers, 2003; Serre, Oliva, et al., 2007). De plus, des études ont montré que l'apprentissage modifiait le traitement des objets visuels (Freedman et al., 2006; Jiang et al., 2007; Op de Beeck, Wagemans, & Vogels, 2003) mais aussi la sélectivité des neurones de la voie visuelle ventrale (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). Ainsi, l'expérience pourrait influencer par différents moyens le traitement des objets que nous voyons tous les jours dans notre environnement.

Par ailleurs, étant donné que IT semble organisé de la même manière chez le singe et l'humain, il doit y avoir, dès la naissance, soit un mécanisme qui permette à la voie visuelle ventrale de s'organiser très rapidement suivant les caractéristiques de l'environnement, soit une structure de base permettant de différencier les objets animés des objets inanimés, capacité très importante pour la survie.

6.2.2. Un rôle pour l'attention dans la catégorisation d'objets?

Les résultats de cette thèse montrent que le rôle de l'attention dans la catégorisation est minime et que ce type de tâche est réalisée de façon automatique puisqu'une différence d'activité entre des images d'animaux et de non-animaux peut être observée en EEG même si les participants sont en train de réaliser une autre tâche (tâche de familiarité dans l'étude 3, tâche centrale de discrimination de lettre dans l'étude de Carlson). Si l'attention n'est pas nécessaire, aurait-elle quand même un rôle?

Une hypothèse est qu'elle pourrait être utile dans les traitements ultérieurs au codage des caractéristiques visuelles. Dans ce cas, elle pourrait agir de deux façons au regard des résultats de cette thèse :

1. Le mécanisme de lecture des caractéristiques dans IT pourrait nécessiter de l'attention. En effet, il n'est pas suffisant que les caractéristiques des objets soient activées sans attention, il faut aussi pouvoir les lire. Les résultats de l'étude 3 montrent que la catégorisation d'objets, et donc le mécanisme de lecture de IT, peut être réalisée en quasi-absence d'attention. Ce type de catégorisation pourrait ne nécessiter que très peu d'attention et puisqu'on demande aux participants de catégoriser les objets présentés en périphérie alors des mécanismes top-down pourraient être mis en place afin de permettre la lecture des caractéristiques en quasi-absence d'attention. Cependant, ces mécanismes top-down ne seraient pas suffisants pour réaliser d'autres types de tâche. Cette hypothèse pourrait expliquer les résultats différents obtenus dans d'autres paradigmes de double tâche, en particulier ceux de M. A. Cohen et al. (2011) discutés dans le chapitre 3.3.2.
2. Un autre processus que l'attention pourrait influencer, sont les processus de mémorisation. Dans l'article 3, nous avons pu montrer que les participants avaient mémorisé à long terme et

implicitement un set d'images présentées dans une tâche de catégorisation animal/non-animal. Même si les tâches de catégorisations peuvent être réalisées en quasi-absence d'attention, les participants pourraient très bien utiliser leur attention pour réaliser la tâche et ainsi traiter en profondeur et mémoriser les images qui leur étaient présentées (en particulier pour les images difficiles). L'attention permettrait alors des traitements ultérieurs dont l'encodage à long terme des images.

L'attention ne servirait donc pas à coder les caractéristiques des objets mais elle pourrait jouer un rôle dans le traitement de ces caractéristiques dans des niveaux ultérieurs à IT.

6.3. Limitation : les caractéristiques des objets... de quoi parle-t-on exactement?

Un des problèmes majeurs des hypothèses utilisées pour expliquer l'ensemble des résultats de cette thèse est la mise en jeu de « caractéristiques » qui seraient codées dans IT. Les caractéristiques d'un objet seraient les informations sur ce qui fait d'un objet cet objet, et les caractéristiques diagnostiques seraient définies par l'interaction entre les demandes de la tâche et l'information disponible (Schyns, 1998). Cependant, nous ne savons pas ce que cela veut dire pour le système nerveux. En fait, le même problème se pose lorsqu'on parle de « représentation ». L'implémentation neurale d'une représentation, ou de caractéristiques dans notre cas, reste toujours mystérieuse.

Cette question est en particulier importante dans les articles 1 et 4 :

- Les résultats de l'article 1 mettent en lumière l'importance de la diversité des stimuli au sein des catégories d'objets utilisées. Pour expliquer cet effet, ainsi que les différences entre les catégorisations d'objets naturels et artificiels, nous émettons l'hypothèse que plus le nombre de caractéristiques est varié au sein d'une catégorie d'objet, plus les performances de catégorisations diminuent. Ainsi, étant donné que la catégorie animale serait plus variée que la catégorie véhicule, les catégorisations d'animaux seraient plus difficiles à réaliser que les catégorisations de véhicules. Ce pourrait être très informatif de savoir exactement ce qu'il en est mais comment faire pour quantifier les caractéristiques similaires et/ou différentes de deux catégories d'objets ? Mohan & Arun, (2012) ont bien développé un modèle à partir de la structure globale d'une image qui pourrait expliquer les résultats de catégorisation, mais qu'est ce qui est vraiment utile et utilisé par le cerveau ?
- Dans l'article 4 on observe un effet d'amorçage plus important dans les essais incongruents lorsque l'amorce et la cible ont beaucoup de caractéristiques communes que lorsqu'elles en ont moins. Ce partage de caractéristiques est basé sur l'idée de l'appartenance à une même catégorie superordonnée dans l'expérience 1 ou de l'appartenance à une même classe d'objet dans l'expérience 2. Qu'en est-il vraiment ? Qu'est ce qui est « partagé » au sein des populations neuronales ?

Différentes possibilités peuvent être envisagées pour définir les caractéristiques en termes de codage au sein des populations neuronales. Une première hypothèse serait qu'elles se réfèrent à certaines propriétés statistiques de l'image comme les fréquences spatiales, contrastes, couleurs ou textures d'une scène visuelle (Oliva & Schyns, 1997, 2000; Schyns & Oliva, 1994). Plus spécifiquement, le spectre d'amplitude d'une scène naturelle pourrait contenir des informations sur la catégorie de l'image et être utilisé par un classifieur pour catégoriser des scènes selon leur contenu (Oliva & Torralba, 2001; Torralba & Oliva, 2003). En fait, les fréquences spatiales et les caractéristiques visuelles diffèrent suivant les catégories, ce qui les rend de fait dissociables (Crouzet & Serre, 2011; Harel & Benin, 2009). Les caractéristiques codées dans la voie visuelle ventrale pourraient être en quelque sorte le résultat de l'activité de neurones répondant aux fréquences spatiales d'une image. Cette hypothèse a surtout été avancée pour expliquer un accès très rapide au « gist » d'une image (Oliva & Torralba, 2006; Oliva, 2005) alors que ce terme est utilisé ici pour les caractéristiques des objets en général mais aussi dans leurs détails.

Une deuxième hypothèse serait que les caractéristiques soient codées comme des conjonctions de traits visuels spécifiques. En effet, j'ai déjà mentionné que plusieurs études avaient proposé que la catégorisation rapide d'objets puisse être basée sur des traits de complexité intermédiaire codés représentés par l'activité des populations de V2 (Walther & Shen, 2014) ou V4 (Girard et al., 2008; Kirchner & Thorpe, 2006). Les caractéristiques pourraient aussi être des configurations plus complexes de traits visuels spécifiques pouvant être traitées automatiquement (Rousselet et al., 2004; Ullman, Vidal-Naquet, & Sali, 2002; Ullman, 2007) grâce à un mécanisme d'intégration des traits codé au sein de la voie ventrale comme proposé par VanRullen (2009).

L'intégration de toutes les configurations en un objet spécifique pourrait ne pas être nécessaire dans des tâches de catégorisations. Les caractéristiques d'un objet seraient alors des configurations de traits visuels familiers. En fait, le système visuel pourrait coder certaines caractéristiques (ou configurations) dans une première vague feed-forward automatique ne nécessitant pas d'attention. Ces caractéristiques pourraient ne pas être reliées à un objet en particulier (pas d'intégration entre les différentes caractéristiques d'un objet) dans cette première étape et c'est seulement grâce à des mécanismes attentionnels que ce lien pourrait avoir lieu. Par exemple si deux animaux, un oiseau et un tigre, sont

présents dans une scène, alors les caractéristiques telles que bec, plumes, fourrure, rayures, pourraient toutes être activées dans une première vague feed-forward. Cependant, les caractéristiques n'appartiendraient pas à un animal en particulier. En fait on ne saurait même pas combien il y a d'animaux dans l'image. Il faudrait attendre que des mécanismes attentionnels puissent intégrer les différentes caractéristiques entre elles suivant leur localisation pour déterminer que l'animal de gauche est un oiseau et celui de droite un tigre.

Cette hypothèse pourrait être testée en comparant les performances de catégorisation pour une image contenant toutes les parties d'un animal mais en une mosaïque mélangée, avec celles pour une image contenant un animal non modifié. Si les performances sont similaires dans ces deux conditions (caractéristiques mélangées ou dans leurs localisations originales) alors cela démontrerait que les caractéristiques de l'animal sont traitées sans être intégrées entre elles en un unique objet et ce de façon robuste (malgré des différences de statistiques de l'image).

6.4. Perspectives

Les hypothèses développées dans cette discussion et faisant suite aux résultats de cette thèse, ouvrent la voie à différentes questions. Je vais ici en décrire trois en relation avec l'article 4 mais bien d'autres questions pourraient être développées. Par exemple il serait très intéressant de pouvoir étudier la mise en place de l'organisation des catégories d'objets dans la voie visuelle ventrale au cours du développement visuel et conceptuel (e.g. en relation avec le développement du langage).

6.4.1. Quelle est la localisation cérébrale de la compétition entre représentations visuelles ?

Dans l'article 4, nous proposons que lorsque deux représentations d'objets sont actives simultanément, elles entreraient en compétition dans la prise de décision et le degré de compétition dépendrait de la similarité entre les deux représentations. Nous pensons que cette compétition s'effectuerait au niveau de IT. Cependant, l'effet d'amorçage dans les temps de réaction pourrait s'expliquer par différents mécanismes. Afin de pouvoir déterminer quelle en est la cause, il serait intéressant d'utiliser le même paradigme d'amorçage couplé avec des enregistrements en IRMf. L'idée serait alors de comparer l'activité cérébrale suivant la relation entre l'amorce et la cible (congruente, incongruente similaire, incongruente éloignée). Cela permettrait de savoir si :

- les mêmes mécanismes cognitifs sont en jeu dans les conditions congruentes et incongruentes. Par exemple en comparant les régions cérébrales activées dans la condition congruente et la condition incongruente.
- la compétition entre les caractéristiques similaires se produit principalement dans IT ou dans le PFC (qui est potentiellement la région qui « lit » IT). En effet, si la résolution du conflit se fait dans IT alors le niveau d'activité cérébrale devrait être plus important dans cette région dans le cas d'une amorce incongruente proche que dans une amorce incongruente éloignée. Si ce n'est pas le cas, alors l'activité cérébrale devrait être similaire dans les deux cas et une autre région cérébrale devrait voir son niveau d'activité augmenter.

Néanmoins, étant donné la résolution spatiale de l'IRMf, cela semble difficile à réaliser.

6.4.2. Quelle est la nature des caractéristiques conduisant à une compétition entre les représentations visuelles ?

La limite la plus flagrante des hypothèses développées dans cette thèse est l'utilisation du terme « caractéristiques » (voir chapitre précédent). Afin de déterminer si les fréquences spatiales (spectre de Fourier) seraient à la base du codage des caractéristiques dans IT, on pourrait imaginer un protocole d'amorçage similaire à celui de l'article 4 mais cette fois en présentant différentes fréquences spatiales comme amorce. Si l'effet d'amorçage observé dans l'article 4 est la conséquence de fréquences spatiales partagées entre l'amorce et l'image cible, alors on devrait observer un effet d'amorçage même si l'amorce ne représente pas un objet spécifique. Le même type d'expérience pourrait être réalisé pour tester l'effet de la présentation de configurations de traits visuels spécifiques sur la catégorisation de la cible. Ainsi, on pourrait essayer de déterminer un peu mieux la nature des caractéristiques visuelles représentées dans la voie ventrale qui sont à l'origine de la catégorisation visuelle.

6.4.3. Représentations visuelles ou/et représentations conceptuelles ?

J'ai mentionné plusieurs fois la possibilité que les représentations codées dans IT pourraient être « seulement » visuelles. En effet, des résultats montrent que les patterns de réponses dans IT (a) codent majoritairement la différence entre objet animé/non-animé et visage/corps (Kiani et al., 2007), (b) sont très similaires pour les humains et les singes (Kriegeskorte et al., 2008), (c) codent mal les catégories subjectives (nous faisons plus de catégories que ce que IT représente) (Mur et al., 2009), (d) codent principalement pour les formes des objets (Baldassi et al., 2013). Cependant, certains résultats suggèrent que les populations de neurones de IT pourraient aussi représenter les propriétés sémantiques des objets visuels (Baldassi et al., 2013; Bell et al., 2008; Carlson et al., 2014).

Il est possible que nous n'ayons pas encore trouvé le bon moyen de tester si les patterns de réponse de IT contiennent seulement des informations visuelles ou contiennent aussi des informations sémantiques plus conceptuelles (appelées aussi multi-modales ou amodales). Le paradigme d'amorçage utilisé dans l'article 4 pourrait être une solution pour répondre à cette question. L'activité cérébrale pour un amorçage visuel image-image et un amorçage plus conceptuel mot-image pourrait être comparée afin de déterminer si l'interférence a lieu dans les mêmes aires cérébrales.

7. Bibliographie

1. Afraz, S.-R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, *442*(7103), 692–695. doi:10.1038/nature04982
2. Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to « building » stimuli: evidence and implications. *Neuron*, *21*(2), 373–383.
3. Andersen, R. A., Hwang, E. J., & Mulliken, G. H. (2010). Cognitive neural prosthetics. *Annual Review of Psychology*, *61*, 169–190, C1–3. doi:10.1146/annurev.psych.093008.100503
4. Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *NeuroImage*, *17*(2), 890–901.
5. Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology*, *87*(6), 3102–3116.
6. Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1013–1026. doi:10.1037/0096-1523.33.5.1013
7. Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469. doi:10.1016/j.visres.2005.01.004
8. Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, *104*(21), 9087–9092. doi:10.1073/pnas.0703300104
9. Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape Similarity, Better than Semantic Membership, Accounts for the Structure of Visual Object Representations in a Population of Monkey Inferotemporal Neurons. *PLoS Comput Biol*, *9*(8), e1003167. doi:10.1371/journal.pcbi.1003167
10. Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., ... Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, *29*(2), 529–535.
11. Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B. H., & Ungerleider, L. G. (2008). Object Representations in the Temporal Cortex of Monkeys and Humans as Revealed by Functional Magnetic Resonance Imaging. *Journal of Neurophysiology*, *101*(2), 688–700. doi:10.1152/jn.90657.2008

12. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2007). Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(7), 1604-1611. doi:10.1093/cercor/bhl071
13. Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, 8(6), 551-565. doi:10.1162/jocn.1996.8.6.551
14. Benton, A. L. (1980). The neuropsychology of facial recognition. *The American Psychologist*, 35(2), 176-186.
15. Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, 303(5919), 696-698.
16. Biran, I., & Coslett, H. B. (2003). Visual agnosia. *Current Neurology and Neuroscience Reports*, 3(6), 508-512.
17. Blankertz, B., Dornhege, G., Schäfer, C., Krepki, R., Kohlmorgen, J., Müller, K.-R., ... Curio, G. (2003). Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 11(2), 127-131. doi:10.1109/TNSRE.2003.814456
18. Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology (2006)*, 61(4), 552-557. doi:10.1080/17470210701798290
19. Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325-14329. doi:10.1073/pnas.0803390105
20. Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area V4. *The Journal of Neuroscience*, 14(2), 554-567.
21. Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination tasks. *Perception & Psychophysics*, 60(1), 1-23.
22. Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics*, 48(1), 45-58.
23. Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7(8), 880-886. doi:10.1038/nn1278
24. Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(2), 369-384.
25. Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 63, 409-412.

26. Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience*, *15*(5), 704–717.
27. Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The Emergence of Semantic Meaning in the Ventral Temporal Pathway. *Journal of Cognitive Neuroscience*, *26*(1), 120–131. doi:10.1162/jocn_a_00458
28. Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10). doi:10.1167/13.10.1
29. Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The Neural Dynamics of Face Detection in the Wild Revealed by MVPA. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(3), 846–854. doi:10.1523/JNEUROSCI.3030-13.2014
30. Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*(2), 225–271.
31. Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, *2*(10), 913–919.
32. Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent Modulation of Category-related Cortical Activity. *Cerebral Cortex*, *12*(5), 545–551. doi:10.1093/cercor/12.5.545
33. Chiu, Y.-C., Esterman, M., Han, Y., Rosen, H., & Yantis, S. (2011). Decoding task-based attentional modulation during face categorization. *Journal of cognitive neuroscience*, *23*(5), 1198–1204.
34. Chun, M. M., & Wolfe, J. M. (2001). Visual Attention. In *Blackwell Handbook of Sensation and Perception* (Vol. Ch 9, p. 272–310). E. Bruce Goldstein.
35. Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*. doi:10.1038/nn.3635
36. Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage*, *32*(2), 583–591. doi:10.1016/j.neuroimage.2006.04.180
37. Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain: A Journal of Neurology*, *123* (Pt 2), 291–307.
38. Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-Scene Perception Requires Attention. *Psychological Science*, *22*(9), 1165–1172. doi:10.1177/0956797611419168
39. Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240–247.

40. Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(8), 2608-2618. doi:10.1523/JNEUROSCI.5547-11.2012
41. Cooper, E. H., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, 68(4), 221-234.
42. Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, 12(1), 125-131.
43. Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) « brain reading »: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, 19(2), 261-270. doi:10.1016/S1053-8119(03)00049-1
44. Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision*, 10(4), 16.1-17. doi:10.1167/10.4.16
45. Crouzet, S. M., & Serre, T. (2011). What are the Visual Features Underlying Rapid Object Recognition? *Frontiers in Psychology*, 2, 326. doi:10.3389/fpsyg.2011.00326
46. Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16(6), 763-770. doi:10.1038/nn.3381
47. D'Lauro, C., Tanaka, J. W., & Curran, T. (2008). The preferred level of face categorization depends on discriminability. *Psychonomic Bulletin & Review*, 15(3), 623-629.
48. Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology*, 32(4), 331-341.
49. De Renzi, E. (1986). Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. *Neuropsychologia*, 24(3), 385-389.
50. Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature neuroscience*, 4(7), 752-758.
51. Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., ... Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, 395(6702), 597-600.
52. Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research*, 40(16), 2187-2200.
53. Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, 1, 21. doi:10.3389/fpsyg.2010.00021

54. Delorme, A., & Thorpe, S. J. (2001). Face identification using one spike per neuron: resistance to image degradations. *Neural Networks: The Official Journal of the International Neural Network Society*, 14(6-7), 795-803.
55. Denkinger, B., & Koutstaal, W. (2009). Perceive-decide-act, perceive-decide-act: How abstract is repetition-related decision learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 742-756. doi:10.1037/a0015263
56. Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 4(8), 2051-2062.
57. DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415-434. doi:10.1016/j.neuron.2012.01.010
58. Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. *Journal of Neuroscience*, 33(4), 1331-1336. doi:10.1523/JNEUROSCI.4081-12.2013
59. Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, 16(10), 1453-1461. doi:10.1093/cercor/bhj086
60. Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)*, 293(5539), 2470-2473. doi:10.1126/science.1063414
61. Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, 96(3), 433.
62. Eckstein, D., & Henson, R. N. (2012). Stimulus/response learning in masked congruency priming of faces: Evidence for covert mental classifications? *The Quarterly Journal of Experimental Psychology*, 65(1), 92-120. doi:10.1080/17470218.2011.590595
63. Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5), 14-14. doi:10.1167/11.5.14
64. Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology*, 26(4), 309-321.
65. Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human LOC carry information about object exemplars within category. *Journal of Cognitive Neuroscience*, 20(2), 356-370. doi:10.1162/jocn.2008.20019
66. Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science (New York, N.Y.)*, 247(4943), 721-723.
67. Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23(1), 115-125.

68. Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601. doi:10.1038/33402
69. Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When Categories Collide: Accumulation of Information About Multiple Categories in Rapid Scene Perception. *Psychological Science*, 22(6), 739-746. doi:10.1177/0956797611407930
70. Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology. Human Perception and Performance*, 31(6), 1476-1492. doi:10.1037/0096-1523.31.6.1476
71. Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience*, 13(2), 171-180.
72. Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, 9(2), 303-308.
73. Fairhall, S. L., Anzellotti, S., Pajtas, P. E., & Caramazza, A. (2011). Concordance between perceptual and categorical repetition effects in the ventral visual stream. *Journal of Neurophysiology*, 106(1), 398-408. doi:10.1152/jn.01138.2010
74. Farah, M. J., McMullen, P. A., & Meyer, M. M. (1991). Can recognition of living things be selectively impaired? *Neuropsychologia*, 29(2), 185-193.
75. Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, 12(6), 893-924. doi:10.1080/13506280444000571
76. Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1), 1-47. doi:10.1093/cercor/1.1.1
77. Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain and Cognition*, 59(2), 145-158. doi:10.1016/j.bandc.2005.06.002
78. Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 16(11), 1631-1644. doi:10.1093/cercor/bhj100
79. Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science (New York, N.Y.)*, 330(6005), 845-851. doi:10.1126/science.1194908
80. Ganaden, R. E., Mullin, C. R., & Steeves, J. K. E. (2013). Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *Journal of Cognitive Neuroscience*, 25(6), 961-968. doi:10.1162/jocn_a_00372

81. Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform « face area » is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, 12(3), 495-504.
82. Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience*, 17(3), 507-517. doi:10.1162/0898929053279522
83. Girard, P., Jouffrais, C., & Kirchner, C. H. (2008). Ultra-rapid categorisation in non-human primates. *Animal Cognition*, 11(3), 485-493. doi:10.1007/s10071-008-0139-2
84. Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20-25.
85. Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137-176. doi:10.1016/j.cogpsych.2008.06.001
86. Greene, M. R., & Oliva, A. (2009b). The briefest of glances The time course of natural scene understanding. *Psychological Science*, 20(4), 464-472.
87. Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, 13(2), 159-166. doi:10.1016/S0959-4388(03)00040-0
88. Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555-562. doi:10.1038/nn1224
89. Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, 3(8), 837-843. doi:10.1038/77754
90. Gross, C. G. (1994). How Inferior Temporal Cortex Became a Visual Area. *Cerebral Cortex*, 4(5), 455-469. doi:10.1093/cercor/4.5.455
91. Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science (New York, N.Y.)*, 166(3910), 1303-1306.
92. Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, 35(1), 96-111.
93. Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *Journal of Experimental Psychology. Human Perception and Performance*, 35(4), 1264-1273. doi:10.1037/a0013621
94. Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10), E962-971. doi:10.1073/pnas.1312567111

-
95. Hasson, U., Hendler, T., Ben Bashat, D., & Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *Journal of Cognitive Neuroscience*, 13(6), 744-753. doi:10.1162/08989290152541412
 96. Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.
 97. Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523-534. doi:10.1038/nrn1931
 98. Hecaen, H., & Angelergues, R. (1962). Agnosia for faces (prosopagnosia). *Archives of Neurology*, 7, 92-100.
 99. Herrnstein, R. J., & Loveland, D. H. (1964). Complex Visual Concept in the Pigeon. *Science*, 146(3643), 549-551. doi:10.1126/science.146.3643.549
 100. Holmes, E. J., & Gross, C. G. (1984). Effects of inferior temporal lesions on discrimination of stimuli differing in orientation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 4(12), 3063-3068.
 101. Horner, A. J., & Henson, R. N. (2011a). Incongruent Abstract Stimulus-Response Bindings Result in Response Interference: fMRI and EEG Evidence from Visual Object Classification Priming. *Journal of Cognitive Neuroscience*, 24(3), 760-773. doi:10.1162/jocn_a_00163
 102. Horner, A. J., & Henson, R. N. (2011b). Stimulus-response bindings code both abstract and specific representations of stimuli: evidence from a classification priming design that reverses multiple levels of response representation. *Memory & Cognition*, 39(8), 1457-1471. doi:10.3758/s13421-011-0118-8
 103. Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science (New York, N.Y.)*, 310(5749), 863-866. doi:10.1126/science.1117593
 104. Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, 76(6), 1210-1224. doi:10.1016/j.neuron.2012.10.014
 105. Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2), 78-87. doi:10.1037/0278-7393.5.2.78
 106. Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 6(1), 1-12.
 107. Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, 12 Suppl 2, 35-51. doi:10.1162/089892900564055

108. Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9379-9384.
109. Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. A. (2013). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*. doi:10.1152/jn.00394.2013
110. Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, 73(1), 218-226.
111. Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, 53(6), 891-903. doi:10.1016/j.neuron.2007.02.015
112. Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7), 499-512. doi:10.1167/3.7.4
113. Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive psychology*, 16(2), 243-275.
114. Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47(26), 3286-3297. doi:10.1016/j.visres.2007.09.013
115. Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 17(11), 4302-4311.
116. Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1476), 2109-2128. doi:10.1098/rstb.2006.1934
117. Kastner, S., & Pinsk, M. A. (2004). Visual attention as a multilevel selection process. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4), 483-500.
118. Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13(1), 90-101.
119. Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, 97(6), 4296-4309. doi:10.1152/jn.00024.2007
120. Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, 46(11), 1762-1776. doi:10.1016/j.visres.2005.10.002
121. Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology. General*, 139(3), 558-578. doi:10.1037/a0019165

-
122. Konorski, J. (1967). *Integrative activity of the brain*. University of Chicago Press: Chicago.
123. Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, 39(2), 184–199.
124. Kovács, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences*, 92(12), 5587–5591.
125. Kreiman, G., Hung, C. P., Kraskov, A., Quiroga, R. Q., Poggio, T., & DiCarlo, J. J. (2006). Object Selectivity of Local Field Potentials and Spikes in the Macaque Inferior Temporal Cortex. *Neuron*, 49(3), 433–445. doi:10.1016/j.neuron.2005.12.019
126. Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, 3(9), 946–953. doi:10.1038/78868
127. Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141. doi:10.1016/j.neuron.2008.10.043
128. Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., ... Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), 5675–5679.
129. Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: concurrent discrimination of form, color, and motion. *Perception & psychophysics*, 61(7), 1241–1255.
130. Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 22(10), 2013–2033.
131. Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596–9601.
132. Li, N., Cox, D. D., Zoccolan, D., & DiCarlo, J. J. (2009). What Response Properties Do Individual Neurons Need to Underlie Position and Clutter « Invariant » Object Recognition? *Journal of Neurophysiology*, 102(1), 360–376. doi:10.1152/jn.90745.2008
133. Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning Shapes the Representation of Behavioral Choice in the Human Brain. *Neuron*, 62(3), 441–452. doi:10.1016/j.neuron.2009.03.016
134. Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, 62(2), 281–290. doi:10.1016/j.neuron.2009.02.025
135. Liu, N., Kriegeskorte, N., Mur, M., Hadj-Bouziane, F., Luh, W.-M., Tootell, R. B. H., & Ungerleider, L. G. (2013). Intrinsic Structure of Visual Exemplar and Category Representations in

- Macaque Brain. *The Journal of Neuroscience*, 33(28), 11346-11360. doi:10.1523/JNEUROSCI.4180-12.2013
136. Liu, Z., & Richmond, B. J. (2000). Response Differences in Monkey TE and Perirhinal Cortex: Stimulus Association Related to Reward Schedules. *Journal of Neurophysiology*, 83(3), 1677-1692.
137. Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology: CB*, 4(5), 401-414.
138. Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552-563.
139. Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7(6), 485-495. doi:10.3758/BF03198265
140. Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The Time-Course of Visual Categorizations: You Spot the Animal Faster than the Bird. *PLoS ONE*, 4(6), e5927. doi:10.1371/journal.pone.0005927
141. Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *European Journal of Neuroscience*, 21(7), 2007-2018. doi:10.1111/j.1460-9568.2005.04029.x
142. Mack, M. L., & Palmeri, T. J. (2011a). The Timing of Visual Object Categorization. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00165
143. Mack, M. L., & Palmeri, T. J. (2011b). Unraveling ultra-rapid categorization. *Journal of Vision*, 11(11), 832-832. doi:10.1167/11.11.832
144. Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 503-535. doi:10.1037/0278-7393.33.3.503
145. Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., ... Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), 8135-8139.
146. Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, 379(6566), 649-652. doi:10.1038/379649a0
147. Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., & Kawano, K. (2005). Population dynamics of face-responsive neurons in the inferior temporal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 15(8), 1103-1112. doi:10.1093/cercor/bhh209
148. McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419-457.

-
149. McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4(4), 310-322. doi:10.1038/nrn1076
150. McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: a face-specific disorder. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 46(1), 1-10.
151. Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, 412(6845), 401-401. doi:10.1038/35086646
152. Miller, B. T., Vytlačil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience*, 23(1), 1-10. doi:10.1162/jocn.2010.21516
153. Missal, M., Vogels, R., Li, C.-Y., & Orban, G. A. (1999). Shape Interactions in Macaque Inferior Temporal Neurons. *Journal of Neurophysiology*, 82(1), 131-142.
154. Missal, M., Vogels, R., & Orban, G. A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cerebral Cortex (New York, N.Y.: 1991)*, 7(8), 758-767.
155. Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science (New York, N.Y.)*, 320(5881), 1355-1359. doi:10.1126/science.1157436
156. Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision*, 12(11), 19-19. doi:10.1167/12.11.19
157. Mruczek, R. E. B., & Sheinberg, D. L. (2012). Stimulus selectivity and response latency in putative inhibitory and excitatory neurons of the primate inferior temporal cortex. *Journal of Neurophysiology*, 108(10), 2725-2736. doi:10.1152/jn.00618.2012
158. Mullin, C. R., & Steeves, J. K. E. (2011). TMS to the lateral occipital cortex disrupts object processing but facilitates scene processing. *Journal of Cognitive Neuroscience*, 23(12), 4174-4184. doi:10.1162/jocn_a_00095
159. Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI--an introductory guide. *Social Cognitive and Affective Neuroscience*, 4(1), 101-109. doi:10.1093/scan/nsn044
160. Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in Psychology*, 4, 128. doi:10.3389/fpsyg.2013.00128
161. Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117(1), 12-22. doi:10.1016/j.bandl.2010.09.013
162. Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 70.

-
163. Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of verbal learning and verbal behavior*, 21(1), 1-20.
164. Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition and popout in visual search. In *The Attentive Brain* (Parasuraman R., p. 279-298). Cambridge: MIT Press.
165. Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, 9(3), 353-383.
166. Nicolelis, M. A. (2001). Actions from thoughts. *Nature*, 409(6818), 403-407. doi:10.1038/35053191
167. Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424-430. doi:10.1016/j.tics.2006.07.005
168. Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In *Cerebral Cortex: Extrastriate Cortex in Primate* (p. 205-242). NEW YORK-PLENUM PRESS.
169. O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17(4), 580-590. doi:10.1162/0898929053467550
170. Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13), 5951-5955.
171. Oliva, A. (2005). Gist of the scene. *Neurobiology of attention*, 696, 64.
172. Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34(1), 72-107. doi:10.1006/cogp.1997.0667
173. Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41(2), 176-210. doi:10.1006/cogp.1999.0728
174. Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3), 145-175. doi:10.1023/A:1011139631724
175. Oliva, A., & Torralba, A. (2006). Chapter 2 Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research* (Vol. 155, p. 23-36). Elsevier. Consulté à l'adresse <http://linkinghub.elsevier.com/retrieve/pii/S0079612306550022>
176. Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences*, 14(1), 22-30. doi:10.1016/j.tics.2009.11.002

-
177. Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *Journal of Neuroscience*, 26(50), 13025-13036. doi:10.1523/JNEUROSCI.2481-06.2006
178. Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews. Neuroscience*, 9(2), 123-135. doi:10.1038/nrn2314
179. Op de Beeck, H. P., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: dimensions can be biased but not differentiated. *Journal of Experimental Psychology. General*, 132(4), 491-511. doi:10.1037/0096-3445.132.4.491
180. Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology*, 68(1), 70-84.
181. Orban, G. A. (2008). Higher Order Visual Processing in Macaque Extrastriate Cortex. *Physiological Reviews*, 88(1), 59-89. doi:10.1152/physrev.00008.2007
182. Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8(7), 315-324. doi:10.1016/j.tics.2004.05.009
183. Palmer, Stephen E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519-526.
184. Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews. Neuroscience*, 8(8), 636-648. doi:10.1038/nrn2195
185. Perani, D., Cappa, S. F., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., & Fazio, F. (1995). Different neural systems for the recognition of animals and man-made tools. *Neuroreport*, 6(12), 1637-1641.
186. Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, 47(3), 329-342.
187. Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. *Journal of Neurophysiology*, 101(5), 2581-2600. doi:10.1152/jn.91198.2008
188. Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology: CB*, 19(4), 319-324. doi:10.1016/j.cub.2009.01.007
189. Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, 2(5), 509-522.
190. Potter, M. C. (2012). Recognition and Memory for Briefly Presented Scenes. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00032

191. Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, 253(5491), 437-438.
192. Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81(1), 10-15.
193. Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: the time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1163-1175.
194. Potter, M. C., Wyble, B., Pandav, R., & Olejarczyk, J. (2010). Picture detection in rapid serial visual presentation: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1486-1494. doi:10.1037/a0018730
195. Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., & Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 19(8), 1806-1819. doi:10.1093/cercor/bhn210
196. Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 16(16), 5205-5215.
197. Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, 1301, 89-99. doi:10.1016/j.brainres.2009.08.085
198. Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research*, 20(1), 37-45. doi:10.1016/j.cogbrainres.2003.11.017
199. Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research*, 46(15), 2336-2343. doi:10.1016/j.visres.2006.01.020
200. Reddy, L., & VanRullen, R. (2007). Spacing affects some but not all visual searches: Implications for theories of attention and crowding. *Journal of Vision*, 7(2), 3-3. doi:10.1167/7.2.3
201. Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, 4(2), 4-4. doi:10.1167/4.2.4
202. Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, 61(2), 168-185. doi:10.1016/j.neuron.2009.01.002
203. Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019-1025. doi:10.1038/14819
204. Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, 136(3), 451-469. doi:10.1037/0096-3445.136.3.451

-
205. Rogers, T. T., Ralph, M. A. L., Hodges, J., & Patterson, K. (2004). Natural selection: the impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology*, 21(2), 331-352. doi:10.1080/02643290342000366
206. Rolls, E. T. (2000). Functions of the Primate Temporal Lobe Cortical Visual Areas in Invariant Visual Object and Face Recognition. *Neuron*, 27(2), 205-218. doi:10.1016/S0896-6273(00)00030-1
207. Rolls, E. T., & Tovéé, M. J. (1994). Processing Speed in the Cerebral Cortex and the Neurophysiology of Visual Masking. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 257(1348), 9-15. doi:10.1098/rspb.1994.0087
208. Rolls, E. T., & Tovéé, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73(2), 713-726.
209. Rolls, E. T., Tovéé, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, 101(3), 473-484.
210. Rosch, E., Mervis, C., Gray, W., Johnson, D., & Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439.
211. Rosinski, R. R. (1977). Picture-Word Interference Is Semantically Based. *Child Development*, 48(2), 643-647. doi:10.2307/1128667
212. Rossion, B., Caldara, R., Seghier, M., Schuller, A.-M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain: A Journal of Neurology*, 126(Pt 11), 2381-2395. doi:10.1093/brain/awg241
213. Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5(7), 629-630. doi:10.1038/nn866
214. Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, 8(8), 363-370. doi:10.1016/j.tics.2004.06.003
215. Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (p. 3-30).
216. Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and Tolerance (« Invariance ») Both Increase as Visual Information Propagates from Cortical Area V4 to IT. *The Journal of Neuroscience*, 30(39), 12978-12995. doi:10.1523/JNEUROSCI.0179-10.2010
217. Sadeh, B., Pitcher, D., Brandman, T., Eisen, A., Thaler, A., & Yovel, G. (2011). Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Current Biology: CB*, 21(22), 1894-1899. doi:10.1016/j.cub.2011.09.030
218. Schiller, P. H. (1995). Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature*, 376(6538), 342-344. doi:10.1038/376342a0

-
219. Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6), 3272-3278.
220. Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(47), 11055-11059. doi:10.1523/JNEUROSCI.2621-05.2005
221. Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, 67(1-2), 147-179. doi:10.1016/S0010-0277(98)00016-X
222. Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, 5(4), 195-200. doi:10.1111/j.1467-9280.1994.tb00500.x
223. Sergent, J., & Signoret, J. L. (1992). Implicit access to knowledge derived from unrecognized faces in prosopagnosia. *Cerebral Cortex (New York, N.Y.: 1991)*, 2(5), 389-400.
224. Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. In T. D. and J. F. K. Paul Cisek (Éd.), *Progress in Brain Research* (Vol. Volume 165, p. 33-56). Elsevier.
225. Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15), 6424-6429.
226. Serre, T., Wolf, L., & Poggio, T. (2005). Object recognition with features inspired by visual cortex. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, p. 994-1000).
227. Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying Object Categories from Event-Related EEG: Toward Decoding of Conceptual Representations. *PLoS ONE*, 5(12), e14465. doi:10.1371/journal.pone.0014465
228. Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. D., & Schacter, D. L. (2003). Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *NeuroImage*, 19(3), 613-626.
229. Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, 35(6), 1157-1165.
230. Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400(6747), 869-873. doi:10.1038/23703
231. Sun, J. Y., & Perona, P. (1996). Preattentive perception of elementary three-dimensional shapes. *Vision Research*, 36(16), 2515-2529.
232. Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology. General*, 130(3), 534-543.

-
233. Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive psychology*, 23(3), 457–482.
234. Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology*, 98(3), 1626–1633. doi:10.1152/jn.00012.2007
235. Thierry, G., Pegna, A. J., Dodds, C., Roberts, M., Basan, S., & Downing, P. (2006). An event-related potential component sensitive to images of the human body. *NeuroImage*, 32(2), 871–879. doi:10.1016/j.neuroimage.2006.03.060
236. Thorpe, S. J. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In G. Eckmiller & G. Hauske (Éd.), *Parallel processing in neural systems* (p. 91–94). Elsevier.
237. Thorpe, S. J., Delorme, A., & VanRullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks: The Official Journal of the International Neural Network Society*, 14(6-7), 715–725.
238. Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking Categories in the Brain. *Science*, 291(5502), 260–263. doi:10.1126/science.1058249
239. Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.
240. Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *The European Journal of Neuroscience*, 14(5), 869–876.
241. Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21(4), 753–759.
242. Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, 63, 483–509. doi:10.1146/annurev-psych-120710-100412
243. Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: computation in neural systems*, 14(3), 391–412.
244. Tovée, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *Journal of Neurophysiology*, 72(3), 1049–1060.
245. Tovée, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, 70(2), 640–654.
246. Treisman, A. (1996). The binding problem. *Current opinion in neurobiology*, 6(2), 171–178.
247. Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97–136.

-
248. Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95(1), 15-48.
249. Treisman, A., & Kanwisher, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. *Current opinion in neurobiology*, 8(2), 218-226.
250. Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14(1), 107-141.
251. Treisman, A., & Souther, J. (1985). Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology. General*, 114(3), 285-310.
252. Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9), 989-995. doi:10.1038/nn1111
253. Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, 105(49), 19514-19519. doi:10.1073/pnas.0809662105
254. Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, 11(2), 58-64. doi:10.1016/j.tics.2006.11.009
255. Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 682-687. doi:10.1038/nn870
256. Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2), 157-165. doi:10.1016/0959-4388(94)90066-3
257. Ungerleider, L. G., & Mishkin, M. (1982). Two Cortical Visual Systems. In *Analysis of Visual Behavior* (The MIT Press., p. 549-586).
258. VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, 17(1-2), 103-119. doi:10.1080/13506280802196451
259. VanRullen, R. (2011). Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00365
260. VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems*, 48(1), 229-239.
261. VanRullen, R., Reddy, L., & Fei-Fei, L. (2005). Binding is a local problem for natural objects and scenes. *Vision Research*, 45(25-26), 3133-3144. doi:10.1016/j.visres.2005.05.012
262. VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience*, 16(1), 4-14. doi:10.1162/089892904322755502
263. VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30(6), 655-668. doi:10.1068/p3029

-
264. VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454-461.
265. VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*(23), 2593-2615.
266. Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81-92. doi:10.1167/5.1.8
267. Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *European Journal of Neuroscience*, *11*(4), 1223-1238. doi:10.1046/j.1460-9568.1999.00530.x
268. Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *European Journal of Neuroscience*, *11*(4), 1239-1255. doi:10.1046/j.1460-9568.1999.00531.x
269. Vogels, R., & Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *12*(7), 756-766.
270. Walther, D. B., & Shen, D. (2014). Nonaccidental Properties Underlie Human Categorization of Complex Natural Scenes. *Psychological science*, 0956797613512662.
271. Wang, Y., Fujita, I., & Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nature Neuroscience*, *3*(8), 807-813. doi:10.1038/77712
272. Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*, *27*(4), 635-657. doi:10.1080/14640747508400525
273. Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. *Brain: A Journal of Neurology*, *106* (Pt 4), 859-878.
274. Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain: A Journal of Neurology*, *107* (Pt 3), 829-854.
275. Warrington, E. K., & Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception*, *7*(6), 695 - 705. doi:10.1068/p070695
276. Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74-97. doi:10.1007/s00426-011-0392-x
277. Weiner, K. S., Sayres, R., Vinberg, J., & Grill-Spector, K. (2010). fMRI-Adaptation and Category Selectivity in Human Ventral Temporal Cortex: Regional Differences Across Time Scales. *Journal of Neurophysiology*, *103*(6), 3349-3365. doi:10.1152/jn.01108.2009

-
278. Weiskrantz, L., & Saunders, R. C. (1984). Impairments of Visual Object Transforms in Monkeys. *Brain*, 107(4), 1033-1072. doi:10.1093/brain/107.4.1033
279. Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238. doi:10.3758/BF03200774
280. Wolfe, J. M. (1998). Visual search. In *Attention* (Vol. 1, p. 13-73). London: University College London Press.
281. Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in cognitive sciences*, 7(2), 70-76.
282. Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics*, 73(6), 1650-1671. doi:10.3758/s13414-011-0153-3
283. Wolfe, J. M., & Bennett, S. C. (1997). Preattentive Object Files: Shapeless Bundles of Basic Features. *Vision Research*, 37(1), 25-43. doi:10.1016/S0042-6989(96)00111-3
284. Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology. Human Perception and Performance*, 15(3), 419-433.
285. Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15(2), 77-84. doi:10.1016/j.tics.2010.12.001
286. Wolpaw, J. R., & McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17849-17854. doi:10.1073/pnas.0403504101
287. Yaginuma, S., Niihara, T., & Iwai, E. (1982). Further evidence on elevated discrimination limens for reduced patterns in monkeys with inferotemporal lesions. *Neuropsychologia*, 20(1), 21-32.
288. Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple Object Response Normalization in Monkey Inferotemporal Cortex. *The Journal of Neuroscience*, 25(36), 8150-8164. doi:10.1523/JNEUROSCI.2058-05.2005
289. Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(45), 12292-12307. doi:10.1523/JNEUROSCI.1897-07.2007

8. Annexes

8.1. Liste des acronymes

BOLD (Blood-Oxygen-Level Dependent) : signal dépendant du niveau d'oxygène sanguin

CGL : corps géniculé latéral du thalamus

EBA (Extrastriate Body Area) : aire extrastriée des corps

EEG : ElectroEncéphaloGraphie

ERP (Event Related Potential) : potentiel évoqué

FFA (Fusiform Face Area) : aire des visages dans le gyrus fusiforme

IRMf : Imagerie par Résonance Magnétique fonctionnelle

IT : cortex Inféro-Temporal

LDA (Linear Discriminant Analysis) : analyse linéaire discriminante

LFP (Local Field Potential) : potentiel de champ local

LOC : cortex occipital latéral

MEG : MagnétoEncéphaloGraphie

MVPA (Multi-Variate Pattern Analysis) : analyse multi-variée de patterns

OFA (Occipital Face Area) : aire des visages dans le cortex occipital inférieur

PDP (Parallel distributed processing) : traitement distribué et en parallèle

PFC (PreFrontal Cortex) : cortex préfrontal

RSVP (Rapid Serial Visual Presentation) : présentation visuelle rapide d'images en série

SOA (Stimulus Onset Asynchrony) : temps entre le début de la présentation d'un stimulus jusqu'au début de la présentation du stimulus suivant

SVM (Support Vector Machines) : Machine à vecteurs de support

TEP : Tomographie par Emission de Positons

TMS (Transcranial magnetic stimulation) : stimulation magnétique transcrânienne

8.2. Poster Mack & Palmeri (2011b)

The Dynamics of Categorization Unraveled

Michael L. Mack
michael.mack@vanderbilt.edu

Thomas J. Palmeri
thomas.palmeri@vanderbilt.edu

catlab.psy.vanderbilt.edu



VANDERBILT UNIVERSITY

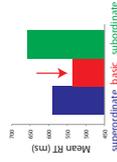
Motivation



We recognize objects at different levels of abstraction.
Some of these category levels are available before others.
Which category level is reached fastest during object perception?

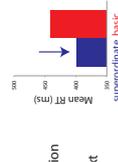
Basic-level Advantage

Standard category verification task



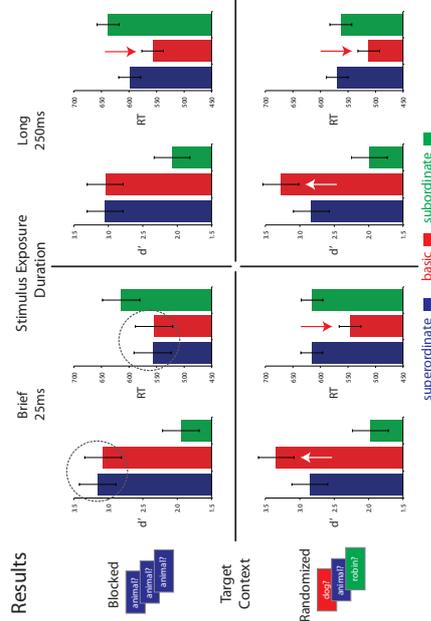
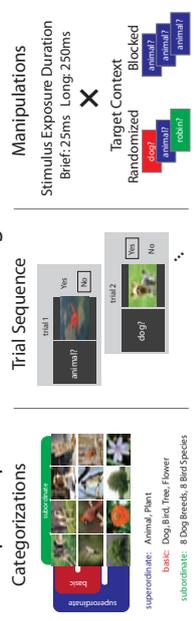
Superordinate-level Advantage

Ultra-rapid category verification task



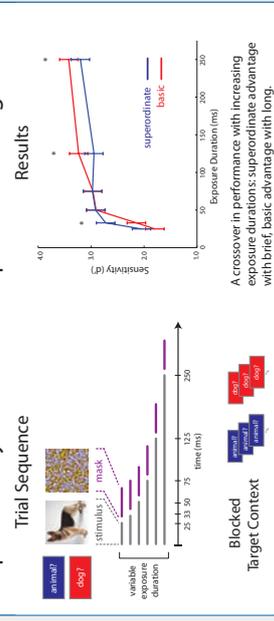
Bridging Paradigms

Exp. 1: Exposure Duration x Target Context

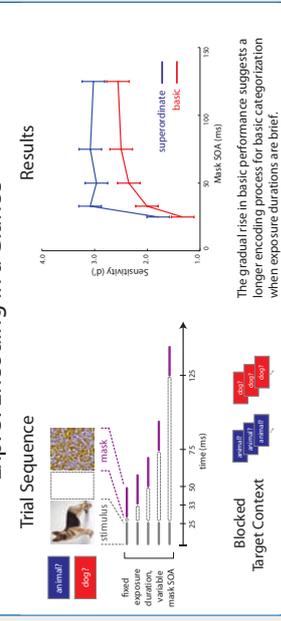


Time Course of Perceptual Encoding

Exp. 2: The Dynamics of Perceptual Processing

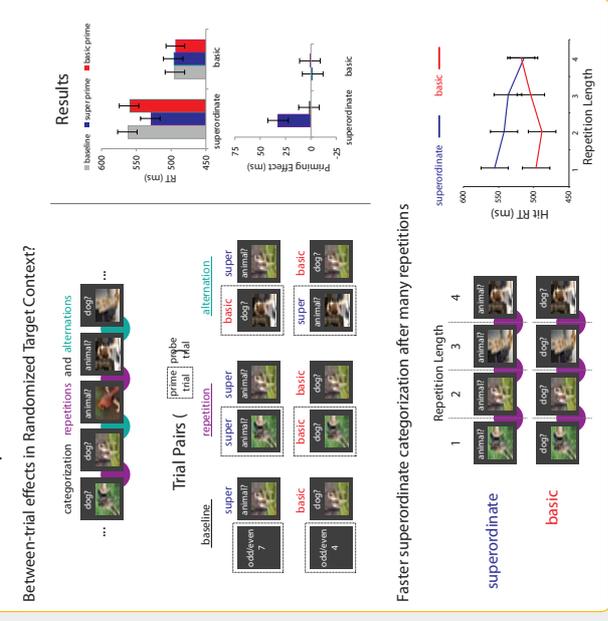


Exp. 3: Encoding in a Glance



Target Category Context

Exp. 4: Local Shifts in Context



Conclusions

The speed of categorization depends on a variety of factors. You may spot animal faster than bird, but only in a glance and when animal is the only thing in mind.

Early in processing, perceptual information for superordinate categories is available quickly leading to a superordinate-level advantage. A basic-level advantage emerges with more time for encoding.

Local shifts in experimental context affect superordinate categorization. Basic categorization is relatively automatic. With only a few repetitions, superordinate categorization is as fast as basic categorization.

References

- Jolicoeur, Gluck, & Kosslyn. (1984). Cognitive Psychology, 16(2).
- Macé, Joubert, Nespoulous, & Fabre-Thorpe. (2009). PLoS ONE, 4.
- Mack & Palmeri. (2010). JEP: HPP, 36(5).
- Mack, Wong, Gauthier, Tanaka, & Palmeri. (2009). Vision Research, 49.
- Rogers & Pattersen. (2007). JEP: General, 136(3).
- Tanaka & Taylor. (1991). Cognitive Psychology, 23.

Acknowledgements

This research funded by the Temporal Dynamics of Learning Center (SBE-0542013) a NSF funded Science of Learning Center, a grant from the James S. McDonnell Foundation, and NEI Grant P30-EY008126.



8.3. Analyses of « no » responses in study 4

Because of the particular categorization tasks that we used in this study, « no » responses are not very informative. We did not ask participants to do a binary categorical decision (such as bird/dog or animal/vehicle) but asked them to report the presence or absence of a particular category (such as bird/non-bird in experiment 1 and animal/non-animal in experiment 2). Hence, they would not be looking for attributes of two categories but for a particular combination of attributes (e.g. bird in Experiment 1) versus an absence of this combination. In this situation, it is hard to determine what exactly the brain would be doing when deciding « no ». It could be using a different combination of attributes for the distractor category or just looking for an absence of target attributes. We therefore decided not to include « no » responses in the main article. Nevertheless, we report these results below.

8.3.1. Experiment 1

We first compared RTs and ERs for « no » responses either for congruent and incongruent trials separately for non-bird animal targets and vehicle targets (Figure 26A). We applied a 2x2x2 repeated measures ANOVA on RTs and ERs using as main factors type of target (non-bird animal or vehicle), congruency (congruent or incongruent), and SOA (80 or 180 ms). The results showed that vehicles were categorized 53 ms faster ($F(1,11)=189.96$, $p=3 \times 10^{-8}$, $\eta^2=0.945$) and with 5% less ERs ($F(1,11)=26.12$, $p=3.10^{-4}$, $\eta^2=0.70$) than animals. As with bird targets, congruent trials were categorized faster ($F(1,11)=37.90$, $p=7 \times 10^{-5}$, $\eta^2=0.775$) and with less ERs ($F(1,11)=12.43$, $p=0.005$, $\eta^2=0.53$) than incongruent trials. We also found an interaction between type of target and SOA ($F(1,11)=19.235$, $p=0.0011$, $\eta^2=0.64$) for RTs but no other effect was significant. Post-hoc analyses showed that the interaction effect was due to a general decrease in RTs (i.e., in both congruent and incongruent trials) at longer SOA for non-bird animal targets (paired t-test $t(11)=2.42$, $p=0.03$) that was absent for vehicle targets (paired t-test $t(11)=0.45$, $p=0.66$).

The priming effect (incongruent – congruent) for RTs and ERs was assessed for both target types separately. In both cases, incongruent trials were trials in which a bird prime preceded the presentation of a non-bird animal or a vehicle. For congruent trials, either two non-bird animals or two vehicles

were presented successively. We applied a 2x2 repeated measures ANOVA (2 types of target x 2 SOAs), which did not find any significant effects. That is, the priming effect was the same for both non-bird animal and vehicle targets.

One could argue that our model predicts a larger priming effect for non-bird animal than for vehicle target after the presentation of a bird prime. In an incongruent trial, the patterns of activity overlap more between a bird prime and a non-bird animal target than between a bird prime and a vehicle target. Thus, excluding the bird representation should take more time for a non-bird animal than for a vehicle target condition. However, participants were overall faster and more accurate in responding « no » to a vehicle than to a non-bird animal. The pattern of activity elicited by a non-bird animal is probably close to the one expected by the system for a bird. In this situation, it is hard for the system to exclude this pattern of activity (or in other words, discard the activity common for a bird and a non-bird animal) compared to a pattern of activity elicited by a vehicle. So the effect of relatedness is present in both congruent and incongruent trials but in this case, it is not evident in the priming effect.

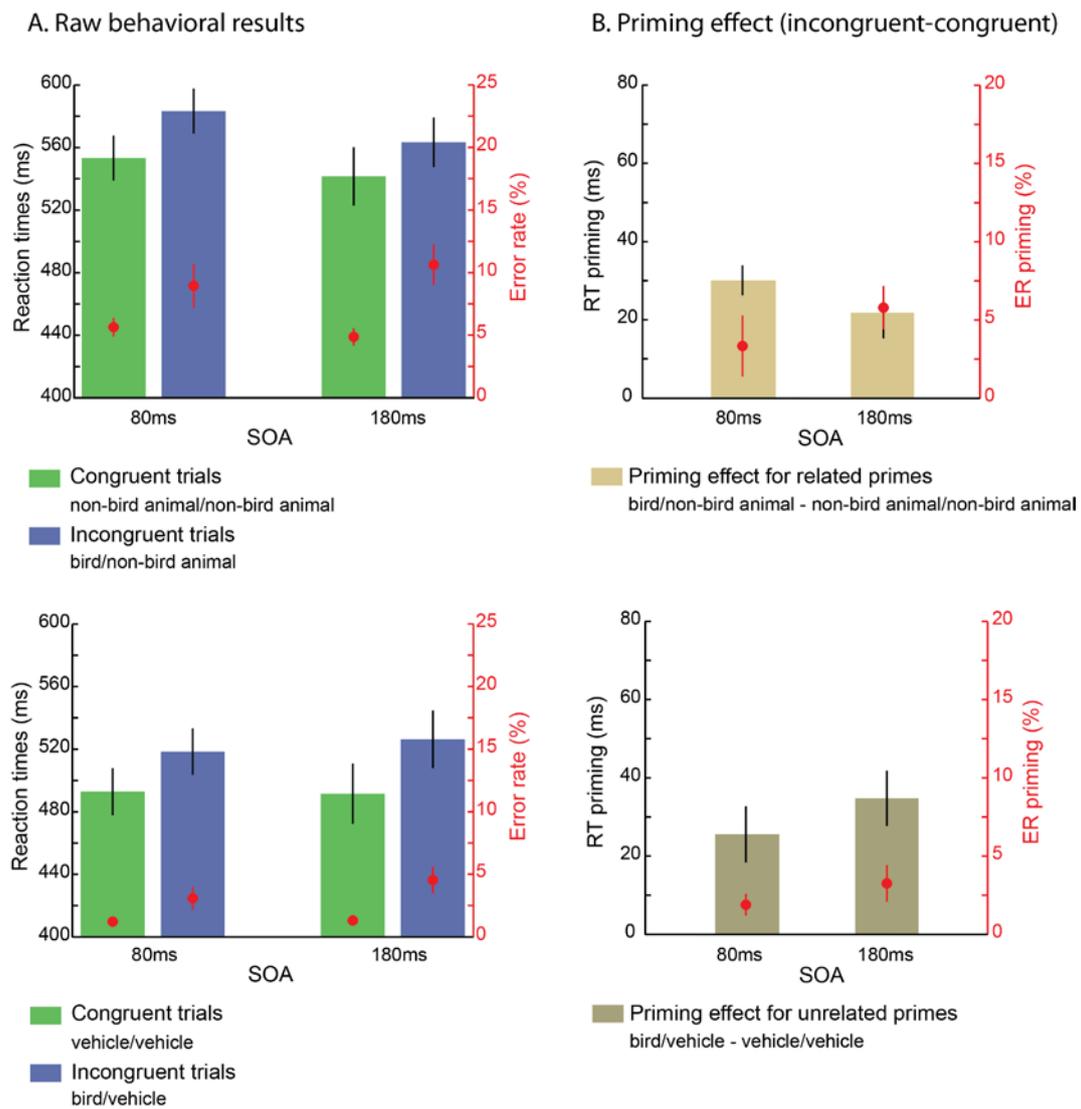


Figure 26. Experiment 1: Performance in bird/non-bird categorization task on distractor images (top row: non-bird animal, bottom row: vehicle). **Panel A: Congruency effect.** Both distractor types were categorized faster (bars) and more accurately (red circles, see right vertical-axis) when they were preceded by a congruent prime (same category, green bars) than an incongruent prime (bird, blue bars) at both 80 ms and 180 ms SOAs. **Panel B: Effect of prime relatedness on the priming effect (difference between incongruent and congruent primes).** At both SOAs ER and RT priming were similar for both type of distractors.

8.3.2. Experiment 2

Analyses similar to that for « yes » responses were conducted for « no » responses in Experiment 2 (animal/non-animal categorization task; Figure 27A). We first applied a 2x2 repeated measures

ANOVA (congruency x SOA) on raw RTs and ERs regardless of the subclass relatedness. Participants were 47 ms faster ($F(1,11)=65.13$, $p=6 \times 10^{-6}$, $\eta^2=0.86$) and 3% more accurate ($F(1,11)=16.58$, $p=0.002$, $\eta^2=0.60$) to categorize vehicle images when they were preceded by vehicle prime (congruent condition) than when they were preceded by an animal prime (incongruent condition). There was no other significant effect.

We then determined the priming effect for related images (prime and target images from the same subclass) and for unrelated images (prime and target images from different subclasses) separately (Figure 27B). We performed a 2x2 repeated measures ANOVA (SOA x relatedness) on these RT and ER priming. ER priming was similar in both conditions; however, RT priming was 11 ms larger for related than unrelated primes ($F(1,11)=8.57$, $p=0.014$, $\eta^2=0.44$) at both SOAs (no effect of SOA or interaction). This effect could be due to two possibilities: (1) vehicles were categorized as « not an animal » faster in the related congruent condition than in the unrelated congruent condition or (2) vehicles were categorized slower in the related incongruent condition than in the unrelated incongruent condition. To test it, we compared raw RTs for related and unrelated primes for congruent and incongruent trials separately. We found that participants categorized vehicle images faster in the related congruent condition than in the unrelated congruent condition (paired t-test $t(11)=3.695$, $p=0.004$) but their RTs were similar for both incongruent related and unrelated conditions (paired t-test $t(11)=0.36$, $p=0.73$).

Thus, the larger RT priming for related primes is mostly explained by participants being faster in the related congruent condition (e.g. air vehicle/air vehicle) compared to the unrelated congruent condition (e.g. air vehicle/ground vehicle). Further, because there was no effect of relatedness in the incongruent condition, the larger RT priming for related primes was not due to shared attributes between animal and vehicles subclasses (conceptual or contextual attributes). Indeed, this effect could be explained by the diversity of stimuli in the each subclass of animals and vehicles (see study 1). Because vehicles that belong to the same subclass share similar shapes (at least more similar shapes than animals from the same subclass), the effect of relatedness in congruent related trials might be related to visual attributes. From the perspective of our model, it then makes sense that it can influence categorization performance at short SOA since these attributes are diagnostic for the task. On the other hand, the effect of relatedness observed for « yes » responses (animal target) was found in both

congruent and incongruent trials but only at long SOA. We therefore argued that it would be the consequence of non-diagnostic attributes (e.g. functionality) that are shared between prime and target (whatever their category, either animal or vehicle).

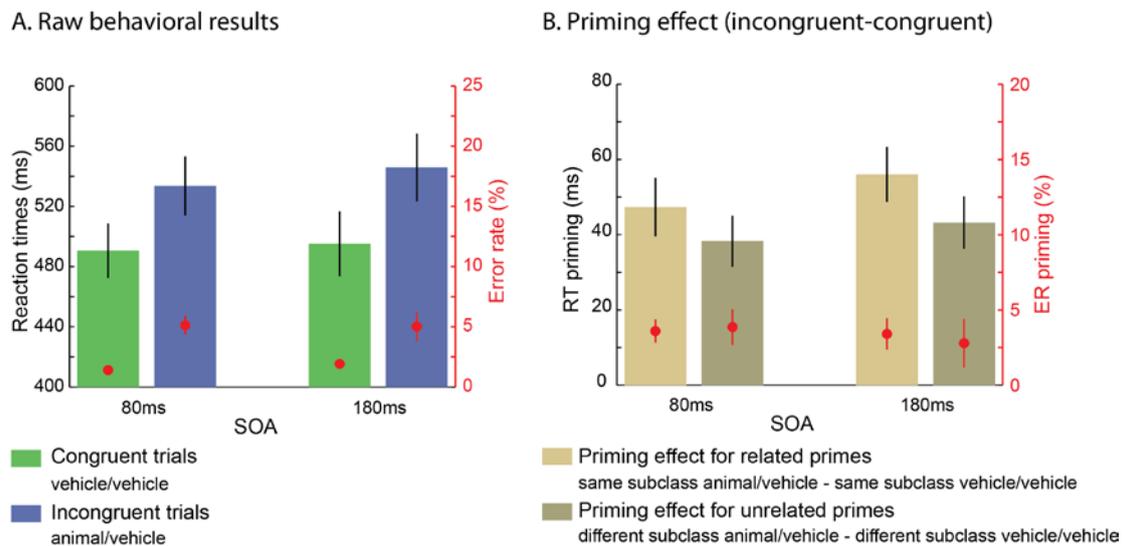


Figure 27. Experiment 2: Performance in animal/non-animal categorization task on distractor (vehicle) images. Each image (prime and target) belonged to one of three subclasses (air, water, ground). Legend as in Figure 2. **Panel A: Priming effect.** Performance was better (shorter RTs and lower ERs) in the congruent than in the incongruent condition. **Panel B: Effect of prime relatedness on the priming effect.** Prime and target were either related (same subclass, e.g. both air objects) or unrelated (different subclasses). There was no effect of relatedness on ER priming. However, RT priming was larger for related than unrelated primes at both SOAs.

8.3.1. Summary

Analyses of « no » responses show the same priming effect as for « yes » responses. That is, participants were faster and more accurate in the congruent than in the incongruent condition. In the first experiment, we also found that participants' performance were lower for rejecting (answering « no ») a non-bird animal than a vehicle. In the second experiment, we found an early effect of relatedness on vehicle categorization but which was only present in the congruent condition and thus could be explained by visual similarity between exemplars from the same subclass. All these results are consistent with the model that we developed.

COGNITIVE MECHANISMS IN VISUAL OBJECT CATEGORIZATION

Categorization is a fundamental process of object recognition. To determine its underlying mechanisms, a series of experiments examined the roles of stimulus properties, categorization level, attention, memory, and category-relatedness in natural scenes categorization tasks. The results of the first study suggest that categorization performance is driven by natural scene characteristics that are diagnostic for the task, and not by other parameters such as stimulus presentation time or response mode. A second study shows that visual objects embedded in complex natural scenes can be categorized in the near-absence of attention. Long-term recognition memory for stimuli encountered in ultra-rapid categorization tasks was tested in a third study. Behavioral and event related potential results suggest that detailed encoding of complex scenes happens without any explicit instruction, and information about object categories is processed automatically. The final study explores the interaction between two concurrently active category representations by presenting two objects (belonging to the same or different categories) in a rapid sequence. The greater the degree of relatedness between two objects, the more affected the processing of the second object is. To explain these results a physiologically inspired model is proposed, which posits that the extent of interaction between concurrently active object categories depends on the extent of overlap between the activity patterns in the infero-temporal cortex elicited by the two objects. The results of this thesis, together with results from other studies, support the idea that visual object characteristics are processed automatically (study 3) in the near-absence of attention (study 2) and represented in the ventral stream in a distributed and continuous manner. Categorization performance would then depend on the similarity between and within the target and the distractor categories (study 1) or on the similarity between two active object representations (study 4).

Keywords: visual categorization, attention, recognition memory, priming, object representation, Infero-temporal cortex (IT)

MECANISMES COGNITIFS DANS LA CATEGORISATION D'OBJETS VISUELS

Auteur : Marlène Poncet

Directeur de thèse : Michèle Fabre-Thorpe

Lieu et date de soutenance : Toulouse, 6 juin 2014

La catégorisation est un processus fondamental de la reconnaissance d'objets. Pour comprendre ses mécanismes sous-jacents, cette thèse interroge dans une série d'expériences le rôle du niveau de catégorisation, de l'attention, de la mémoire, et de la relation entre les catégories d'objets, dans la catégorisation de scènes naturelles. Les résultats de la première étude indiquent que les performances de catégorisation sont influencées par les caractéristiques diagnostiques de la tâche, mais pas par d'autres paramètres tels que le temps de présentation des stimuli ou le mode de réponse. Une seconde étude montre que des objets visuels présentés dans des scènes naturelles complexes peuvent être catégorisés en quasi-absence d'attention. La mémoire de reconnaissance à long-terme pour des stimuli rencontrés dans des tâches de catégorisations ultra-rapides a été testée dans une troisième étude. Les résultats en comportement et en potentiels évoqués indiquent que les scènes complexes sont encodées sans instruction explicite et catégorisées automatiquement. La dernière étude explore les interactions entre deux représentations d'objets actives simultanément en présentant deux objets l'un après l'autre dans une période de temps très courte. Plus le degré de relation entre deux objets est élevé, plus le traitement du second objet est affecté. Pour expliquer ces résultats, un modèle inspiré de la physiologie est proposé, qui postule que le niveau d'interaction entre des catégories d'objets actives simultanément dépend du niveau de chevauchement entre les patterns d'activité du cortex inféro-temporal produits par chacun des objets. Les résultats de cette thèse, en relation avec ceux d'autres études, sont compatibles avec l'idée que les caractéristiques visuelles des objets sont traitées automatiquement (étude 3) en quasi-absence d'attention (étude 2) et représentées dans la voie visuelle ventrale de façon distribuée et continue. Les performances de catégorisation dépendraient alors de la similarité des catégories cibles et distracteurs (étude 1) ou de la similarité entre les représentations actives de deux objets (étude 4).

Mots-clés : catégorisation visuelle, attention, mémoire de reconnaissance, amorçage, représentation d'objet, cortex inféro-temporal

Discipline : Neurosciences

Laboratoire : CNRS CERCO UMR 5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex