Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Cotutelle internationale avec :

Présentée et soutenue par :

Maxime Cauchoix

Le 3 décembre 2012

Titre :

Etude sur la pensée animale:
Continuité neuro-cognitive de la catégorisation visuelle

École doctorale et discipline ou spécialité :

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

Centre de Recherche Cerveau et Cognition

Directeur(s) de Thèse :

Denis Fize (CerCo, Toulouse)
Michèle Fabre-Thorpe (CerCo, Toulouse)

Rapporteurs :

Odile Petit (IPHC, Strasbourg)
Olivier Pascalis (LPNC, Grenoble)

Autre(s) membre(s) du jury :

Thomas Serre (Brown University, Providence, USA)
Jean-Michel Lassalle (CRCA, Toulouse)

REMERCIEMENTS

A Odile Petit et Olivier Pascalis qui ont eu le courage d'accepter l'évaluation de ce manuscrit.

A Jean Michel Lassalle et Thomas Serre pour leur participation au jury de thèse.

A Gladys, Michèle et Catherine qui ont eu la lourde tâche de trouver et corriger mes innombrables fotes d'orthographes (s'il en reste, c'est la leur … de faute ;-).

A Michèle pour m'avoir motivé, encouragé et accompagné dans l'écriture de ce manuscrit.

A Thomas pour m'avoir accueilli dans son laboratoire et pour m'avoir autant aidé quand j'en avais besoin. Aussi bien pour les analyses pointues, pour l'écriture des articles que pour surmonter cette épreuve initiatique, qu'est la thèse.

A Seb pour les discussions théoriques (à différent niveaux) et cette coopération parfaite dans l'écriture des articles.

A Denis pour m'avoir supporté (dans tous les sens du terme) pendant ces 4 ans.

A Dicky, Roucky et Rox, les trois macaques, qui ont fait de leur mieux pour nous aider à les étudier, et à tous ceux qui ont contribué à l'amélioration de leur bien être : Michèle, Luc, Franck, Angeline, Camille, Gregory.

A Romain, pour les 400 coups (il y en avait peut être plus).

A mon père, pour m'avoir éveillé à la magie du monde animal et de la nature.

A ma mère, pour son énergie, son écoute et sa patience.

A mon grand père Jean, pour m'avoir donné le gout de la science.

A ma grand mère Catherine, pour m'avoir donné celui de la connaissance.

A ma grand mère Yvette, pour m'avoir appris à relativiser toutes ces choses.

A Gladys, pour ce qu'elle m'apporte à chaque instant.

La forme la plus élevée de la vie animale est la girafe.

Le nerf optique est celui qui amène les idées lumineuses au cerveau.

Jean Charles

A tous les animaux qui ont souffert de la curiosité humaine,

PREFACE

L'intrigue

Il est intrigant, lorsqu'on a à peine 10 ans, et qu'on se retrouve collé de curiosité à l'épaisse vitre de l'enclos des chimpanzés du zoo de Beauval, dans le centre de la France, d'observer une vieille femelle chimpanzé se diriger gentiment vers son abreuvoir, s'emplir abondamment la bouche d'une eau à la couleur douteuse, retourner se percher sur le faux rocher le plus haut de l'enclos et, par un geste d'une extrême précision, émettre un puissant jet d'eau qui s'envole au dessus de la vitre pour vous retomber en plein dans la figure. Cette humide anecdote pose assez bien l'intrigue de cette thèse, dans toute la variété de sens que la langue française prodigue à ce mot:

- intrigue, comme provenant du verbe intriguer signifiant: qui excite vivement la curiosité. Le monde animal recèle des comportements aussi surprenants que celui-ci qui ont poussés l'homme, de tous temps, à chercher à savoir ce qui se passe dans la tête des animaux.
- intrigue, comme une machination secrète ou déloyale pour nuire à quelqu'un. Sans nul doute, la machination était déloyale, d'une nuisance modérée mais totalement secrète. Le secret pour moi est à deux niveaux: celui du monde mental de cette vieille guenon et celui des processus neuronaux à l'origine du monde mental de cette vieille guenon.
- intrigue, comme une succession de faits et d'actions qui forment la trame d'un roman, d'un film ou d'une pièce de théâtre. Dans le cas présent, le film est une observation de terrain, et les faits et actions sont un enchaînement de comportements, eux-mêmes sous-tendus par une cascade de processus cognitifs qu'il appartient à l'éthologue cognitiviste de décortiquer dans les moindres détails.

Qu'est qui pousse à faire une thèse?

Sûrement à cause de cette aventure, et de bien d'autres encore, avec les différents chevaux, chiens, chats, rongeurs et corvidés qui ont égaillé mon enfance, mon souhait le plus fort à cette époque, était d'être capable, tel Yakari, de lire dans les pensées des animaux. C'est drôle comme certains rêves de gosse sont tenaces.

La vie nous offre chaque jour, à travers les rencontres, les voyages, internet, des

dizaines de nouvelles opportunités. Des milliers de chemins se dessinent devant nous. On s'y essaie, on en parcourt les premiers kilomètres et bien souvent on s'y perd. Pourtant au milieu de ce fouillis incompréhensible, certains repères nous montrent la direction à prendre, certaines idées nous guident et nous font avancer. Plus que des idées, ce sont des croyances intimes. Des convictions profondes qui vous relèvent quand le poids du doute vous écrase, qui vous font franchir les obstacles les plus insurmontables ou qui simplement, donnent un sens à la vie. Pour beaucoup il s'agit de dieux à adorer, pour nous, les esprits involontairement scientifiques, ce sont des théories à explorer, à démontrer...

Naturellement, ces idées vous poussent à commencer une thèse, mais surtout, ce sont elles qui vous permettent d'en venir à bout.

Ma conviction c'est que les animaux pensent, à leur manière, mais ils pensent.

Un rêve de gosse? une candide évidence? un grossier anthropomorphisme? ou un problème métaphysique insoluble?

Prenons le temps d'une thèse, quatre ans de réflexions, de recherches bibliographiques et d'expériences scientifiques, pour triturer la question.

Parce que la pluralité des points de vue est une richesse et parce que mon coeur balance entre ces deux domaines, le sujet sera abordé à la fois à travers une approche comportementale et neuroscientifique. Au cours de ce mémoire de thèse, nous essaierons de montrer comment les avancées techniques et idéologiques récentes dans ces deux disciplines éclairent ce débat sur la pensée animale.

PLAN

1 INTRODUCTION GÉNÉRALE 10

1.1 DE QUOI PARLE CETTE THÈSE ? 10
1.2 LA PENSÉE ANIMALE : HISTOIRE D'UN SUJET CONTROVERSÉ 12
1.2.1 NAISSANCE D'UNE CONTROVERSE (XVII AU XVIIIÈME) 12
1.2.2 UNE THÉORIE POUR EXPLIQUER LE VIVANT QUI S'APPLIQUE AUSSI AUX COMPORTEMENTS ET À LA PENSÉE (XIXÈME) 14
1.2.3 BEHAVIOURISME ET ETHOLOGIE : LE RETOUR À L'ANIMAL-MACHINE (DÉBUT DU XXÈME SIÈCLE JUSQU'AUX ANNÉES 70) 16
1.2.4 UN MONDE MENTAL POUR L'ANIMAL : ÉTHOLOGIE COGNITIVE ET PSYCHOLOGIE COMPARÉE (FIN DES ANNÉES 1970 JUSQU'À LA FIN DES ANNÉES 2000) 19
1.2.5 LA CONTINUITÉ COGNITIVE HOMME-ANIMAL REMISE EN CAUSE : DÉBATS CONTEMPORAINS 22
1.3 PROBLÉMATIQUE GÉNÉRALE: UNE DÉFINITION OPÉRATIONNELLE DE LA PENSÉE POUR L'ÉTUDIER CHEZ L'ANIMAL 24
1.4 LES CONCEPTS COMME BRIQUES DE LA PENSÉE 25
1.5 UN MONDE DE CATÉGORIES 26

2 COGNITION COMPARÉE DE L'ABSTRACTION VISUELLE : 35

2.1 INTRODUCTION 35
2.1.1 LES 4 QUESTIONS DE TINBERGEN : UN CADRE THÉORIQUE POUR ÉTUDIER LE COMPORTEMENT ANIMAL 35
2.1.2 LA COGNITION COMPARÉE : DÉFINITION, ENJEUX ET LIMITES 37
2.1.3 LES MÉTHODES POUR ÉTUDIER L'ABSTRACTION CHEZ L'ANIMAL 41
2.1.4 L'ABSTRACTION VISUELLE: UNE FONCTION LARGEMENT RÉPANDUE AU SEIN DU RÈGNE ANIMAL 48
2.1.5 COGNITION ET CONCEPT CHEZ LES PRIMATES 56
2.2 ARTICLE1 : HUMAINS ET MACAQUES RHÉSUS PARTAGENT DES REPRÉSENTATIONS VISUELLES HAUT NIVEAU 67
69
2.2.1 RÉSUMÉ DE L'ARTICLE 70
2.3 DISCUSSION 74
2.3.1 CATÉGORISATION RAPIDE DE SCÈNES NATURELLES : UNE « VRAIE» CATÉGORISATION? 74
2.3.2 POURQUOI LES SINGES PEUVENT-ILS ÊTRE INFLUENCÉS PAR UNE ASSOCIATION OBJET/CONTEXTE ? 75
2.3.3 QU'EST CE QUE LES ANIMAUX COMPRENNENT DANS UNE IMAGE ? 77
2.3.4 CONCEPT NATURELS : PLUS LOIN DANS L'ABSTRACTION ? 81
2.3.5 CONCEPTS ABSTRAITS BASÉS SUR DES RÈGLES 82

3 NEUROSCIENCES COMPARÉES DE L'ABSTRACTION VISUELLE 85

3.1 ENTRÉE EN MATIÈRE 85
3.2 LE SYSTÈME VISUEL 87
3.3 MATÉRIEL ET MÉTHODES : COMMENT ÉTUDIER LES BASES BIOLOGIQUES DES REPRÉSENTATIONS MENTALES ? 91
3.3.1 COMMENT MESURER L'ACTIVITÉ CÉRÉBRALE ? 92
3.3.2 NOUVELLES MÉTHODES D'ANALYSE POUR DÉCODER LES REPRÉSENTATIONS MENTALES 97
3.3.3 ARTICLE 2 : L'ANALYSE MULTI-VARIÉE APPLIQUÉE AUX ENREGISTREMENTS INTRACRÂNIENS CHEZ LE SINGE 105
3.4 CODAGE NEURONAL DES OBJETS ET CATÉGORIES CHEZ LE SINGE 114
3.4.1 INTRODUCTION 114

3.4.2 ARTICLE 3 : DES REPRÉSENTATIONS PRÉCOCES DANS LES AIRES VISUELLES INTERMÉDIAIRES PERMETTENT LA CATÉGORISATION RAPIDE 119
3.4.3 DISCUSSION 124
3.5 CODAGE NEURONAL DES OBJETS ET CATÉGORIES CHEZ L'HOMME 125
3.5.1 ENTRÉE EN MATIÈRE 125
3.5.2 TRAITEMENT DES OBJETS ET CATÉGORIES NATURELLES DANS LA VOIE VISUELLE VENTRALE CHEZ L'HOMME : APPORT DES MÉTHODES HÉMODYNAMIQUES (IRMF, PET) 126
3.5.3 TRAITEMENT CÉRÉBRAL RAPIDE DE CATÉGORIES NATURELLES : APPORTS DE L'EEG DE SURFACE 131
3.5.4 PROBLÉMATIQUE GÉNÉRALE 133
3.6 ARTICLE 4 : DÉCODAGE DE LA CATÉGORIE DE VISAGE EN EEG DE SURFACE 135
3.7 A LA RECHERCHE D'UNE HOMOLOGIE NEUROPHYSIOLOGIQUE FONCTIONNELLE DE LA CATÉGORISATION RAPIDE 143
3.8 DISCUSSION 155
3.8.1 ARTICLE 5: SUR LA POSSIBILITÉ D'UN TRAITEMENT PUREMENT ASCENDANT DE L'INFORMATION VISUELLE DANS LA VOIE VENTRALE 155
3.8.2 ARTICLE 6: SUR LA POSSIBILITÉ D'UNE ORIGINE SOUS-CORTICALE DE LA VISION RAPIDE 159

4 DISCUSSION GÉNÉRALE 169

4.1 RÉSUMÉ DES RÉSULTATS 169
4.2 SUR LA NÉCESSITÉ D'UNE APPROCHE COMPARATIVE MÉTICULEUSE 171
4.3 UNE PENSÉE CHEZ L'ANIMAL ? RÉPONSE À LA PROBLÉMATIQUE GÉNÉRALE 179
4.4 OUVERTURE ET PERSPECTIVES 181

5 ANNEXE 187

5.1 « THE ONLY REAL VALUABLE THING IS INTUITION ». ALBERT EINSTEIN 187

1 Introduction générale

1.1 De quoi parle cette thèse ?

La question qui a déclenché l'existence de cette thèse paraît simpliste ou inabordable par une approche scientifique rigoureuse: est-ce que les animaux pensent? Pourtant, nous verrons dans l'introduction comment elle anime les débats entre grands penseurs depuis plusieurs siècles. On présentera deux lignes de pensée: (i) celle engendrée par Descartes qui traite les animaux comme des machines dénuées de toute capacité psychique contrairement à nous les humains doués de raison et (ii) celle dont le principal instigateur est Darwin, qui prône une continuité du vivant impliquant l'existence d'une pensée animale dont la différence avec celle de l'homme serait de degrés et non de nature.

Nous verrons ensuite comment le courant actuel du cognitivisme fournit une définition opérationnelle pour l'étude de la pensée (animale): celle de représentation mentale. Nous expliquerons comment la notion de représentation mentale et celle de concept sont interdépendantes et synonymiques. Le terme fourre-tout et abstrait de pensée, peut alors prendre un sens bien précis et abordable par les méthodes scientifiques actuelles : la capacité à former et à manipuler des concepts (représentation mentale de catégories d'objets ou d'événements existant dans le monde extérieur). Nous focaliserons alors notre investigation de la pensée animale sur l'étude des concepts, de leur formation et de leur base neurophysiologique. Nous choisirons en particulier l'étude des concepts visuels qui bénéficient, de loin, de la littérature la plus riche et la plus rigoureuse.

Dans un premier chapitre résumant la littérature disponible, nous montrerons à quel point les capacités de catégorisation et de reconnaissance, témoignant de la possession de tels concepts, sont répandues au sein du règne animal. A tel point que nous serons amenés à questionner l'existence d'une réelle évolution et phylogénie de la cognition au profit de systèmes cognitifs propres à chaque espèce, résultant de pressions environnementales similaires et d'adaptations convergentes. La clé permettant de lever cette ambiguïté résidant dans la démonstration de fonctions cognitives homologues supportées par un substrat biologique commun, c'est à la quête de ce

Graal que le reste de cette thèse sera dédié. En particulier, nous nous intéresserons au macaque rhésus (macaca mulatta) et à l'homme (homo sapiens), deux espèces de primates phylogéniquement proches, et à leur capacité à former des catégories naturelles abstraites. Nous utiliserons pour cela la tâche de catégorisation rapide dans laquelle des images naturelles sont flashées au sujet. Ce dernier doit alors répondre le plus rapidement possible en relâchant un bouton si l'image contient un animal.

Nous commencerons par une étude comportementale dans laquelle nous montrerons qu'homme et macaque, dans ce protocole, présentent des réponses très similaires qui suggèrent l'existence d'une base commune aux concepts visuels que possèdent ces deux primates. Si les singes sont capables d'apprendre des concepts visuels humains, cela ne nous renseigne pas sur leur niveau de compréhension et d'interprétation de ces concepts. En utilisant des images d'un prédateur ancestral, les serpents, nous montrerons que, dans cette tâche, les singes sont bien capables d'extraire le sens des images présentées et d'adapter leur comportement, quitte à aller à l'encontre du protocole de renforcement.

Nous plongerons ensuite dans la matière cérébrale, d'abord celle du macaque. Dans un premier temps, nous expliquerons comment la combinaison de nouvelles techniques d'enregistrement multi-électrodes et de techniques d'analyse multivariées des signaux électrophysiologiques permet d'accéder aux contenus des représentations mentales. On est ainsi capable de prédire, essai par essai, quel type de concept visuel est formé dans le cerveau du singe. Nous nous arrêterons notamment sur l'implication des aires intermédiaires de la voie visuelle ventrale pour réaliser la tâche de catégorisation rapide, en montrant un lien étroit entre performances comportementales et activité neurale précoce dans cette voie.

Dans un second temps, nous montrerons que ces méthodes d'analyse développées pour tenter de lire les concepts chez le singe, permettent aussi de lire les représentations mentales de sujets humains, suggérant l'existence d'un codage neuronal commun à ces deux primates. On peut ainsi lire très précocement les concepts visuels associés à une tâche de catégorisation rapide en utilisant le potentiel enregistré en EEG chez des sujets sains. Enfin, en engageant des patients épileptiques (implantés avec des électrodes intracrâniennes pour raisons médicales) dans la même

tâche que celle réalisée par les singes, nous rapporterons des résultats préliminaires suggérant un substrat et des mécanismes neuronaux communs à l'homme et au macaque.

Les différents résultats et concepts constituant cette thèse seront introduits et discutés, au fur et à mesure de leur présentation, de la manière la plus objective possible. Néanmoins, dans un dernier chapitre, nous nous attarderons à quelques spéculations et propositions, empreintes de subjectivité.

1.2 La pensée animale : histoire d'un sujet controversé

La question de la pensée animale a toujours fasciné l'homme. Chaque peuple, chaque civilisation, chaque religion possède sa propre conception du sujet. Il est fort probable que les premiers hommes préhistoriques, confrontés plus que quiconque aux animaux, avaient déjà un avis assez tranché sur la question. Les Egyptiens, plus que d'attribuer une psyché à l'animal, en avaient fait des dieux à adorer. Aristote (384-322 av. J.-C.), en Grèce, s'intéressait déjà aux comportements et à l'intelligence « pratique » des animaux, notamment mis en place pour la reproduction ou le soin aux jeunes. Mais pour lui, seul l'homme était doué du logos (capacité à utiliser une langue étendue à la rationalité).

Il serait extrêmement intéressant de dresser un historique complet des conceptions sur la mentalité animale, mais dans ce mémoire de thèse nous nous limiterons aux courants qui influencent directement nos travaux. Nous verrons que depuis le XVIIème siècle se sont alternativement succédés les défenseurs de l'animal-machine et de l'animal-savant. Ce débat déchaîne toujours les passions contemporaines. Il est encore à l'origine de nombreux articles scientifiques, d'essais philosophiques et sa forte exposition médiatique, à travers tous les moyens de communication actuels, reflète l'intérêt du grand public pour le sujet.

1.2.1 Naissance d'une controverse (XVII au XVIIIème)

S'il est indéniablement le père fondateur de toutes les sciences modernes, Descartes a aussi marqué de son empreinte la conception de l'animal pour les siècles qui lui ont succédé. Sa théorie de l'animal-machine, publiée dans la cinquième partie du

Discours de la méthode et la lettre au marquis de Newcastle, compare l'animal à un pur automate auquel il n'accorde strictement aucune pensée ou autre activité mentale. Seul l'homme est doué de raison. Comme le fonctionnement de la machine, celui de l'animal repose sur un système de tuyauterie. Les seules différences qui les séparent étant:

- la taille des tuyaux. Alors que dans la machine l'homme ne voit qu'eux, dans l'animal ils sont si petits qu'ils ne peuvent être perçus.
- la complexité des mécanismes en jeux. L'animal est certes une machine mais à tout jamais inaccessible à l'ingéniosité humaine.

La thèse de l'animal-machine va déchaîner les passions. Au XVIIème siècle, Malebranche, cartésien convaincu, propose des arguments de nature diverse pour l'étayer. Pour lui, si les animaux peuvent démontrer une forme d'intelligence, c'est parce qu'ils sont réglés comme une montre ou une plante. Il y a intelligence là où il n'y a pas de hasard. Comme pour la montre, l'intelligence de l'animal renvoie surtout à celle de son concepteur : Dieu. L'animal mange sans plaisir et crie sans douleur.

La critique de la thèse de l'animal-machine ne se fait pas attendre et elle est virulente. François Bernier fait appel au bon sens : personne ne pourra jamais croire qu'un animal qu'on écorche vif, qui crie et se débat ne possède pas des sensations bien supérieures à celles d'un morceau de parchemin qu'on déchire. Réaumur dans un élan anticartésien place certaines capacités de l'animal au-dessus de l'être humain. Ainsi dans *Mémoire pour servir à l'histoire des insectes,* il dresse un tableau élogieux des abeilles et de leur capacité de construction. Ces considérations de Réaumur ont particulièrement fait réagir Buffon pour qui l'animal n'innove pas, n'a pas de mémoire, n'a ni vertu ni société, et dont la pensée comme le langage sont inexistants. Même s'il reconnaît des « forces intérieures » aux animaux, il les pense néanmoins emprisonnés par leur sensorialité, au contraire de l'homme qui montre son activité mentale par l'usage de signes. Pour lui, même les singes manquent complètement d'intériorité. « … et en supposant qu'il n'eut que des pensées de singe, il parlerait aux autres singes ; mais on ne les a jamais vus s'entretenir ou discourir ensemble. Ils n'ont donc pas même un ordre, une suite de pensées à leur façon, bien loin d'en avoir de semblables aux nôtres … ».

Locke, en 1700, cherche le compromis ; s'il refuse des « idées complexes » et une capacité d'abstraction à l'animal, il admet qu'il puisse posséder des sentiments et un raisonnement sur des idées « simples ». Leibniz, en 1714, accorde même à l'animal une certaine raison. Celle de l'homme en diffère car elle est plus réflexive et lui permet d'avoir conscience de son existence. Charles-Georges Le Roy est le premier à rapporter une description précise des comportements animaux dans leur milieu, dans *Lettres sur les animaux.* Pour lui, l'étude de l'animal éclaire celle de l'homme car tous deux sentent, agissent en fonction de cette sensibilité, et disposent d'un langage de l'action. Il va même jusqu'à proposer un langage animal articulé.

Au XVIIème et début du XVIIIème, l'opposition entre l'animal et l'homme passionne et les points de vue s'affrontent à l'extrême. Les penseurs de l'époque se battent à coup d'arguments de bon sens, d'idées philosophiques et de principes théologiques. Darwin lui, va faire entrer ces questions dans le champ de la science.

1.2.2 Une théorie pour expliquer le vivant qui s'applique aussi aux comportements et à la pensée (XIXème)

Si Linné et Cuvier font entrer l'homme dans la classification du vivant, Lamarck est le premier à proposer une évolution du singe vers l'homme. Le terrain est alors prêt pour permettre à Darwin de développer sa théorie de l'évolution par la sélection naturelle.

Charles Darwin a provoqué un véritable tremblement de terre dans la façon d'étudier et d'appréhender le vivant. A une époque où les biologistes sont avant tout des naturalistes qui passent le plus clair de leur temps à découvrir, décrire et recenser de nouvelles espèces, Darwin va proposer pour la première fois un mécanisme expliquant à la fois la diversité du vivant et son origine: la sélection naturelle. Si les voyages forment la jeunesse, ils forment aussi les grandes théories dans le cas de Charles Darwin. C'est au cours de 5 années d'un périple naturaliste à travers le monde à bord du HMS Beagle que le jeune Charles (alors âgé de 25 ans) a pu constater l'extraordinaire diversité du vivant. Il fut particulièrement marqué par les îles Galápagos, au large de l'équateur, dont chacune semblait posséder sa propre espèce d'iguane, de tortue et de pinson. Ce n'est que 22 ans plus tard qu'il publiera dans son

livre « On the origin of species » (1859) sa théorie révolutionnaire. Dans ce livre, Darwin expose une théorie selon laquelle, étant donné que tous les individus d'une espèce diffèrent au moins légèrement, certains individus seront plus adaptés à leur environnement, et par conséquent, se reproduiront plus facilement. Ainsi, comme les individus sélectionnés transmettent leurs caractères à leur descendance, les espèces évoluent et s'adaptent en permanence à leur environnement. C'est ce qu'il appelle la "sélection naturelle". Il ne manquait plus à cette théorie unificatrice des sciences du vivant qu'un substrat. Celui-ci, l'ADN et en particulier les gènes, ne furent découverts qu'au cours du siècle dernier. Les gènes expliquent à la fois la variabilité des individus au sein d'une espèce par des mutations aléatoires, et l'hérédité par la transmission du patrimoine génétique. L'intégration de la sélection naturelle de Darwin et de la génétique a permis de créer la théorie synthétique de l'évolution ou néodarwinisme tel que nous la connaissons aujourd'hui. Néanmoins, dans son premier livre Darwin ne fait pas allusion aux origines de l'homme. Il ne le fera que 10 ans plus tard, en 1871, en publiant *The Descent of Man, and Selection in Relation to Sex.* Dans ce livre il formule une réponse aux idées spiritualistes d'Alfred Russel Wallace, pourtant co-découvreur de la théorie de la sélection naturelle. Ce dernier pensait que l'esprit humain était trop complexe pour avoir évolué graduellement, et proposait, en contradiction avec ses propres convictions évolutionnistes, que la conscience et les facultés cognitives supérieures de l'homme étaient le fruit de "the unseen universe of Spirit". Dans *The Descent of Man* et *The Expression of the Emotions in Man and Animals*, paru un an plus tard, Darwin explique que chacune des capacités dites supérieures, présentes chez l'homme, peut être retrouvée dans le règne animal. La différence entre les facultés mentales de l'homme et de l'animal n'est pas de nature mais de degré, en accord avec le perfectionnement graduel de ses capacités provenant d'un ancêtre commun, par la sélection naturelle. Les preuves de Darwin sur les facultés cognitives des animaux sont très empiriques et clairement anthropomorphiques. Néanmoins, l'idée d'une différence de degré et non de nature a été, et reste encore, le fer de lance de la psychologie/cognition comparée. Le travail de tous les scientifiques qui ont appartenu ou qui appartiennent à ces champs de recherche consiste à tenter de montrer avec rigueur ce que Darwin a imaginé et supposé à partir de quelques observations naturalistes. J'aimerais souligner trois idées que m'inspire l'œuvre de Darwin. La première est que la plus grande théorie de la biologie émane de l'approche naturaliste et de l'observation de terrain. La seconde est

qu'il aura fallu 22 ans à Darwin pour publier ses idées. Et la dernière est que Darwin en proposant la continuité des processus cognitifs entre les espèces, est le père officiel de l'approche comparée défendue ici.

A l'instar de Darwin, Thomas Huxley et Ernst Haeckel affirment que l'origine de l'homme est animale. Le premier décrit le singe comme un brouillon de l'être humain et son cerveau, comme le plan de ce qui a été soigneusement effectué dans celui de l'homme. Le second évoque pour la première fois la notion d'erreur anthropocentrique (considérer l'homme comme au centre et but du fonctionnement du monde) en analogie à l'erreur géocentrique (considérer la Terre comme centre de l'Univers). Broca en s'appuyant sur l'anatomie défend lui aussi l'appartenance de l'homme au règne animal. Il met en évidence de nombreux organes communs suggérant la proximité de l'un et de l'autre.

Malheureusement, ces idées évolutionnistes de la fin du 19ème siècle vont rapidement laisser place à des thèses racistes, notamment par la recherche du chaînon manquant: homme inférieur, lien entre l'animal et l'homme occidental. Büchner dresse ainsi une liste de critères qui selon lui différencient les hommes "vrais" des hommes inférieurs: le mariage, la pudeur, la science des nombres, l'usage de vêtements, le suicide, l'agriculture, etc.

A la même époque, l'engouement pour la soudaine possibilité d'une intelligence animale va mener certains auteurs à attribuer hâtivement et sans preuves scientifiques des capacités humaines aux animaux. Ainsi, Georges Romanes en publiant "Animal intelligence" en 1982 (premier traité de psychologie comparée) est accusé de soutenir des propos anthropomorphiques, lorsqu'il dit que les animaux sont doués de jalousie ou de raisonnement.

1.2.3 Behaviourisme et Ethologie : le retour à l'animal-machine (Début du XXème siècle jusqu'aux années 70)

La rigueur du behaviourisme : l'école américaine

Les thèses anthropomorphiques de Romanes vont susciter la révolte de ses contemporains, notamment de Conway Llyod Morgan qui dans son "Introduction to

Comparative Psychology" appelle à l'objectivité dans l'interprétation des comportements. Ce sont les prémices du behaviourisme. Mais c'est à John Watson qu'on attribue généralement la fondation de cette approche sceptique du comportement. Avec lui, on assiste à une sorte de retour à l'animal-machine de Descartes. Pour lutter contre la psychologie introspective, trop peu fiable à leur goût, les behaviouristes considèrent le cerveau et la pensée comme une boîte noire. Pour eux, seul le comportement (behaviour en anglais) induit par des stimuli, mérite d'être étudié. Une thèse, reprise par Frederick Skinner, qui invente le conditionnement opérant, dans lequel l'animal se conditionne lui-même à travers l'utilisation d'un dispositif ingénieux, qu'on appellera vite la cage de Skinner. Dans une telle cage, l'animal-automate associe une récompense à la réalisation d'un comportement dirigé par un stimulus (presser un levier, picorer une clef). Il n'aura alors de cesse de répéter ce comportement, machinalement. Pendant tout ce début du XXème siècle et jusque dans les années 60, l'école américaine voit l'animal comme une machine à associer un stimulus avec une réponse comportementale, apprenant par une loi d'essai/erreur. On lui interdit donc la possession d'une quelconque pensée ou flexibilité comportementale. Par dessus tout, on ne conçoit pas comment étudier un tel phénomène de manière rigoureuse et scientifique.

L'éthologie objectiviste, la sociobiologie et l'écologie comportementale: l'animal esclave de ses gènes

Pendant ce temps-là, en Europe, on voit naître un courant scientifique qui, au lieu d'amener les animaux au laboratoire, se concentre sur l'étude de leurs comportements dans leurs milieux de vie naturels. Cette éthologie objectiviste est portée, entre autres, par Konrad Lorenz, Niko Tinbergen et Karl Von Frisch, tous trois nobélisés en 1973. L'éthologie aporte une attention toute particulière à la notion d'instinct: connaissance façonnée par l'évolution propre à chaque espèce. Par cet intérêt pour les connaissances innées qui guident les comportements, l'éthologie s'oppose aux méthodes expérimentales de conditionnement et à l'apprentissage du behaviorisme. Cependant, les deux domaines s'accordent sur la limitation du monde mental animal. Le déterminisme génétique des instincts défendu par les éthologues n'accorde pas plus de capacités réflexives à l'animal que les associations stimuli/réponses des behaviouristes.

Dans les années 60, on s'aperçoit que la vision de l'éthologie sur l'acquisition d'un répertoire comportemental par la sélection au niveau de l'espèce ou du groupe, ne suffit pas à expliquer l'ensemble des comportements observés dans la nature. En particulier, on se demande comment l'altruisme, comme on peut l'observer dans les sociétés d'insectes où des individus asexués participent à l'élevage d'une descendance qui n'est pas la leur, peut se développer par l'intermédiaire de la sélection naturelle. Préoccupé par ces limitations, William Hamilton propose, en 1964, une théorie de l'évolution génétique des comportements sociaux et fonde par la même occasion un nouveau domaine: la sociobiologie. Hamilton introduit la sélection de parentèle. Les individus stériles d'une colonie d'insectes, en aidant leur mère à se reproduire, contribuent à la propagation de leur génome. Il en va de même pour certains types d'oiseaux chez qui les jeunes de l'année passée, aident les parents à élever leurs frères et sœurs de l'année. La sélection de parentèle explique, dans le cadre de la sélection naturelle, l'altruisme envers un apparenté.

Dans les années 80, les nouveaux concepts apportés par la sociobiologie vont rendre obsolète l'éthologie objectiviste de Lorenz et Tinbergen au profit d'une nouvelle discipline : l'écologie comportementale. Cette discipline, comme la sociobiologie, veut rendre compte des variabilités comportementales intra-spécifiques mais dans un champ plus large que celui limité aux relations sociales. C'est Krebs et Davies qui introduisent le nom et les grands principes de cette science dans leur article « An introduction to behavioural ecology » publié en 1981. Cette discipline considère l'individu comme un *animal economicus* qui gère des budgets temps-énergie afin d'optimiser les bénéfices (survie, énergie, santé, descendance, etc.) et de minimiser les coûts. Face aux contraintes du milieu, tous les individus n'ont pas les mêmes chances de s'adapter et différentes stratégies seront mises en place. L'écologie comportementale a remis au goût du jour la variabilité inter-individuelle, idée clé de la théorie de Darwin, dans l'étude du comportement.

Les trois disciplines présentées ici appartiennent au domaine plus général de la biologie évolutionniste, dont le but est d'expliquer comment un fait biologique actuel (*i.e.* un comportement) est le résultat de l'évolution par sélection naturelle. Cette approche considère toujours l'animal comme une sorte d'automate dont les actes et réponses aux stimuli environnementaux sont prédéterminés, façonnés par des milliers

d'années d'évolution, et dont chaque comportement est orienté vers un but unique, celui de la dissémination de son patrimoine génétique. En ce qui concerne l'éthologie comportementale, il faut noter que cette science a beaucoup évolué depuis sa création et elle rejoint actuellement les sciences cognitives sur de nombreuses facettes.

1.2.4 Un monde mental pour l'animal : éthologie cognitive et psychologie comparée (fin des années 1970 jusqu'à la fin des années 2000)

Le courant cognitiviste, émergeant après la seconde guerre mondiale, s'oppose radicalement au behaviorisme: à l'instar de Bartlett, c'est justement « à ce qui se passe dans la boîte noire » que ses détenteurs s'intéressent. C'est la pensée, le monde mental des individus, qui fait l'objet d'étude des sciences cognitives. Le cerveau est vu comme un système de traitement de l'information, capable d'acquérir, de transformer, d'utiliser, de conserver et de transmettre des connaissances (« cognitio » en latin). Il existe entre l'objet extérieur et le comportement, une *représentation mentale* de l'objet et une représentation mentale du comportement. Ce courant est, par défaut, pluridisciplinaire: il naît des récents progrès en informatique, qui cherchent à créer un « esprit artificiel » (en particulier du champ de la cybernétique), et des neurosciences, qui découvrent le fonctionnement des neurones. Il est rapidement adopté par la psychologie cognitive qui l'utilise comme nouveau cadre de travail. Enfin, il bénéficie des avancées conceptuelles récentes en linguistique et philosophie de l'esprit.

Sous le joug du behaviourisme et de l'éthologie objectiviste, les sciences du comportement animal tardent à se rallier à l'engouement général, et ce n'est que vers la fin des années 70 que l'on commencera à parler d'une cognition chez l'animal. En 1976, Donald Griffin fait explicitement le lien entre ces deux domaines en créant le champ de l'éthologie cognitive. Griffin réagit aux interdits posés par le behaviourisme. Il estime que de nombreux comportements des animaux suggèrent que ces derniers forment des images mentales des objets, des événements et des relations et qu'ils possèdent des capacités de conscience de soi, d'intentionnalité, et d'anticipation des actions futures. Ses livres sur la pensée et la conscience animale remettent au goût du jour les lubies des premiers suiveurs de Darwin tel que Romanes. Néanmoins, cette fois, Griffin s'appuie sur des concepts solides, venus des sciences cognitives, tout en gardant l'approche écologique et évolutionniste de

l'éthologie. Griffin estime que la peur de l'anthropomorphisme n'est pas pertinente. En s'appuyant sur les théories évolutionnistes, il fait remarquer que conscience et pensée sont des outils formidables pour survivre dans l'environnement et qui ont donc tout lieu d'exister dans les espèces animales contemporaines. En prenant les éthologues objectivistes, encore réticents à une cognition de l'animal, à leur propre jeux du darwinisme, il invoque le fait que la continuité neuronale observée entre l'homme et l'animal ne peut être totalement décorrélée d'une continuité de la fonction.

Dans les années 80, un philosophe, Daniel Dennett tente de comprendre les changements conceptuels apportés par l'éthologie cognitive. Il discute les présupposés de l'étude de l'esprit animal. Pour Dennett, même les discours behaviouristes les plus rigoureux sont empreints d'intentionnalité ; il est donc préférable d'assumer pleinement notre propension à attribuer des états attentionnels aux animaux (anthropomorphisme) pour les juger au plus juste. On entre dans une aire complètement nouvelle pour l'étude de l'animal. C'est l'hypothèse de travail qui change. Au lieu d'étudier l'animal en partant du présupposé qu'il possède des capacités minimales, on présuppose des capacités maximales pour ensuite éventuellement les réfuter. Cette approche est valable puisqu'elle est extrêmement féconde. A l'heure où le behaviourisme et l'éthologie classique s'essoufflent, l'envolée de la psychologie cognitive chez l'homme fournit un réservoir immense de thématiques de recherche pour l'éthologie cognitive. Cette approche, ainsi légitimée théoriquement par un philosophe, convainc petit à petit collègues et étudiants, qui ont continué à développer et à porter le champ de l'éthologie cognitive jusqu'à nos jours. On peut citer Bekoff, qui, par ses travaux sur le jeu chez les canidés, défend l'existence d'émotions et de sentiments chez l'animal, ou encore Cheney et Seyfarth, qui, dans des expériences de terrain, ont mis en évidence l'existence de cris sémantiques indiquant le type de prédateur chez le singe vervet. Les thématiques de recherche de l'éthologie cognitive se penchent uniquement sur les processus mentaux impliqués dans les comportements écologiques ou naturels de l'animal, tels que le choix du partenaire sexuel, la recherche de nourriture, la reconnaissance de ses propres petits, etc.

Au contraire, la psychologie dite comparée expérimentale, se veut plus généraliste et vise à identifier les mécanismes cognitifs élémentaires, plus abstraits, communs à travers les espèces, tels que la catégorisation, la résolution de problème, l'attention, la planification, les différentes formes de mémoire ou encore l'orientation spatiale. Dans les années 1980, les travaux d'O'Keefe et Nadel mettent en évidence l'existence biologique d'une « carte cognitive » dans l'hippocampe. Un concept qu'avait déjà imaginé Tolman dans les années 1930, suite à l'observation de rats capables d'utiliser des raccourcis, jamais empruntés auparavant, pour se déplacer dans un labyrinthe expérimental. Cette preuve biologique d'une représentation mentale légitime son utilisation pour la psychologie comparée et devient la notion clef autour de laquelle cette science se construit. Pour les initiateurs du domaine, à l'instar de Roiblat (1982) et Gallistel (1990), la représentation est une correspondance entre l'environnement et le comportement, permettant à l'animal de s'adapter au milieu. Elle permet à tout animal d'analyser les relations spatiales, temporelles ou sociales entre différents objets de l'environnement et ainsi d'anticiper les conséquences de ses actions. Cette vision du fonctionnement de la pensée a aussi été qualifiée de *computationnelle* puisqu'on imagine les comportements animaux (et ceux de l'homme) comme soutenus par différentes *opérations* sur ces représentations.

Le concept de représentation ne fait pas référence à la conscience, c'est un moyen d'objectiver l'étude de la pensée animale. Il donne à la pensée une définition opérationnelle qui permet aux psychologues comparatistes, comme Terrace (1984), d'aborder la question de la pensée animale avec les outils de la psychologie expérimentale, sans se soucier de la subjectivité animale. En ce sens, la psychologie comparée diffère dramatiquement de l'éthologie cognitive de Griffin qui d'ailleurs, n'a jamais utilisé directement l'idée de représentation. Une autre différence fondamentale entre ces deux domaines est que la psychologie comparée travaille exclusivement sur l'animal en laboratoire en utilisant les méthodes du behaviourisme. Néanmoins, il ne faut pas confondre behaviourisme et psychologie comparée car leurs buts sont opposés. La psychologie comparée amène l'animal à répondre dans un dispositif expérimental, comme dans le conditionnement opérant, mais dans le seul but d'interroger la nature des représentations de celui-ci. L'étape fondamentalement différente entre une expérience de psychologie comparée et une expérience behaviouriste, est la phase de transfert. En effet, pour montrer que l'animal n'est pas

coincé dans une simple association stimulus-réponse, le psychologue comparatiste utilise cette phase de transfert, dans laquelle il teste la capacité de l'animal à réagir à de nouveaux stimuli ou à une nouvelle situation. Une telle flexibilité témoigne de l'existence de représentations mentales. Enfin, la finalité avouée de la psychologie comparée, est clairement de comprendre l'évolution de la cognition humaine alors que celle des éthologues cognitivistes est de tester l'hypothèse gradualiste de Darwin sur la continuité entre la cognition de l'homme et l'animal. Ces diverses différences ont tendance à s'atténuer avec le temps et il est actuellement difficile de faire la différence ente un éthologue cognitiviste et un psychologue comparatiste, l'ensemble des méthodes et objectifs de ces deux domaines pouvant être regroupé dans un champ plus large qui est celui de la cognition comparée dont nous parlerons plus longuement dans le chapitre 2.

Pour résumer, depuis les années 1970 jusqu'à nos jours, une littérature abondante, provenant de plusieurs courants, soutient l'existence d'une pensée animale riche, comparable à la nôtre par beaucoup d'aspects. Un consensus semble s'être établi autour de l'hypothèse gradualiste (ou de continuité) de Darwin, qui prône une différence de degré et non de nature entre la cognition animale et humaine.

1.2.5 La continuité cognitive homme-animal remise en cause : débats contemporains

L'éthologie cognitive et la psychologie comparée ont bien fait leur travail. L'image scientifique moderne de l'animal, capable d'utiliser des outils, prévoyant pour le futur, ayant une conscience de soi ou complotant pour fixer des stratégies politiques, semblait avoir comblé –peut-être dans un consensus mou- le fossé cognitif creusé par Descartes et ses successeurs, qui le séparait de l'être humain. Darwin avait raison. Mais voilà que récemment, trois articles sont venus ébranler la confiance en l'hypothèse de continuité cognitive homme-animal qui s'était lentement installée. Ces trois articles présentent chacun une théorie ou des données expérimentales prouvant, selon eux, l'existence d'une discontinuité cognitive entre l'homme et l'animal.

L'équipe de Tomasello, au Max Plank Institute de Lepzig en Allemagne, défend depuis toujours l'importance de la cognition sociale chez l'être humain. Dans leur papier publié dans la revue prestigieuse Science, en 2007 (Herrmann et al. 2007), à

travers une batterie de tests cognitifs menés chez des enfants de 2,5 ans, des chimpanzés et des orang-outans, ces chercheurs défendent l'hypothèse d'une intelligence culturelle spécifique à l'être humain. Alors que les deux espèces de grands singes obtiennent des performances équivalentes à celles des enfants sur la plupart des tests de cognition physique (permanence de l'objet, utilisation d'outils, jugement de quantité ,…), ils sont significativement moins performants sur les tests de cognition sociale (apprentissage social, compréhension d'intentions, …). Pour les auteurs, ce serait donc l'existence de capacités sociales spécifiques à l'homme, apparaissant tôt dans son développement et lui permettant d'échanger et partager des connaissances, qui le rendrait unique dans le règne animal.

Le second article écrit par David Premack (Premack 2007), recense plusieurs études ayant révélé des capacités cognitives de haut niveau chez l'animal, préalablement considérées comme propres à l'homme, telles que: l'éducation, la planification, le raisonnement, le langage, etc. Le constat qu'il dresse est que: oui, les animaux peuvent posséder des compétences cognitives exceptionnelles, mais celles-ci ne sont que des adaptations pour servir un but précis. La capacité de mémoire épisodique chez le geai ne lui sert qu'à retrouver ses caches de nourriture. La spécificité de l'homme est d'avoir des compétences générales qui peuvent servir plusieurs buts. L'animal est spécialisé alors que l'homme est généraliste.

Enfin, le dernier article de l'équipe de Povinelli se veut ostensiblement provocateur dès son titre : « L'erreur de Darwin : explication de la discontinuité entre l'esprit humain et non humain » (Penn et al. 2008). Les auteurs tissent leur message à travers une large revue de la littérature. Pour eux seul l'humain est capable de « réinterpréter » le monde en termes de forces causales et d'états mentaux invisibles. En particulier, les animaux seraient incapables de rejouer mentalement des concepts relationnels. Il leur manquerait la capacité à raisonner analogiquement.

A la lueur de cet historique, on réalise à quel point la pensée animale a déchaîné et déchaîne encore les passions. L'homme a besoin de l'animal pour se définir. C'est pourquoi les questions du propre de l'homme et de la continuité avec l'animal resteront toujours des sujets bouillants, faisant le beurre des philosophes et des médias. Après des siècles de points de vue extrêmes, on croyait la question endormie

dans un consensus général, émanant de la croyance sans faille des biologistes contemporains dans les conceptions darwiniennes. Mais le débat n'est pas terminé comme le montrent ces théories récentes : il reste à creuser, à découvrir à prouver. Il y a matière à faire une thèse.

1.3 Problématique générale: une définition opérationnelle de la pensée pour l'étudier chez l'animal

Pour résumer et simplifier les conceptions sur la nature de l'esprit, il semblerait que la plupart des philosophes et penseurs s'accordent sur deux notions : celles de conscience et de représentation.

La conscience est subjective, c'est le ressenti, elle est donc propre à chaque espèce et même à chaque individu. Cet aspect de la pensée, comme l'explique Thomas Nagel dans son célèbre article « what is it like to be a bat? », est donc par définition inaccessible à autrui et ne fournit pas un moyen opérationnel d'étudier la pensée animale.

La deuxième notion, défendue par les cognitivistes, postule qu'avoir un esprit c'est être capable de former des représentations. Cette définition est opérationnelle car les représentations peuvent être étudiées, c'est l'enjeu des sciences cognitives (Proust 1997). Elles peuvent être caractérisées par la psychologie comparée, modélisées par l'intelligence artificielle et leur existence biologique mise en évidence par les neurosciences. Dans ce cadre, la question "est-ce que les animaux ont une pensée?" peut donc être abordée en résolvant les points suivants :

- est-ce que la psychologie comparée peut nous aider à mettre en évidence un système représentationnel chez l'animal?
- est-ce que ce phénomène est modélisable par les outils de l'intelligence artificielle?
- les neurosciences nous permettent-elles d'en révéler l'existence biologique et d'en découvrir l'origine neurale?

Notre référentiel absolu en matière d'être vivant doué de pensée est notre propre espèce, l'humain. Pour appuyer notre démonstration, il serait donc extrêmement

convainquant de montrer que les outils développés pour prouver l'existence d'une pensée chez l'animal, en répondant aux trois questions précédentes, s'appliquent aussi chez l'homme. On en conclura alors que, selon une définition représentationnelle de la pensée, non seulement les animaux pensent, mais qu'en plus, nos pensées partagent une origine commune avec les leurs.

1.4 Les concepts comme briques de la pensée

Quand on pense, on ne fait pas que rejouer ce que l'on vient d'entendre ou de voir. On est plutôt comme traversé par des idées qui nous permettent de donner un sens à ce monde. En permanence, on classifie, on trie, on associe les différents percepts qui émanent de nos sens. On les organise d'une manière reproductible et intelligible. On les utilise pour spéculer sur le futur. On les combine pour raisonner sur des événements que nous n'avons jamais directement vécus. En résumé, sans nous en rendre compte, nous sommes perpétuellement en train de conceptualiser le monde qui nous entoure. Selon Medin, les concepts et les catégories seraient les briques de la pensée et des comportements humains (Medin 1989).

La capacité à former des concepts a été, et est encore, fortement associée à la capacité langagière de notre espèce. Selon le philosophe anglo-saxon John Locke, il est clair que seuls les hommes possèdent le niveau d'abstraction nécessaire pour former des concepts qui transcendent la perception de l'instant présent (Locke 1690).

Néanmoins, comme nous l'avons vu précédemment, la théorie de Darwin prône une continuité du vivant et une origine commune des traits physiques et mentaux entre homme et animal, qui, logiquement, devraient s'appliquer aussi à cette capacité à conceptualiser le monde. Comme il existe une variabilité, essai par essai, inhérente à tout acte perceptif, il semble évident qu'une espèce ne possédant pas une fonction, même rudimentaire, pour extraire les invariants et organiser les objets et les événements en catégories, serait éteinte depuis longtemps (Ashby & Lee 1993).

La controverse actuelle sur la capacité des animaux à conceptualiser vient certainement de la définition de ce qu'est un concept et du niveau d'abstraction qu'il sous-entend. Par exemple, si on définit généralement les concepts comme les représentions mentales qui nous permettent de répondre de manière similaire à

différents objets en les plaçant dans une même catégorie ; mais pour Geach, pour être de « vrais concepts » ils doivent, en plus, remplir une autre condition: on doit être capable d'agir mentalement sur eux pour découvrir et comprendre de nouvelles propriétés du monde (Geach 1957). Ce n'est pas parce que l'on a entraîné un chien à discriminer des triangles de carrés, qu'il est capable d'inférer les lois de la géométrie. Geach dans la lignée de Locke n'autorise que l'espèce humaine à posséder de vrais concepts.

Mais si les concepts sont uniquement humains, d'où viennent-ils et comment sont-ils apparus? Comme le soutient Wasserman (2008), aux lueurs de la théorie Darwinienne, aucun scientifique actuel ne peut concevoir l'existence d'un attribut de l'espèce humaine qui n'aurait pas d'origine phylogénétique. Et, si les concepts humains sont uniques, on devrait être capable de mettre en évidence des capacités non-conceptuelles chez les animaux chez qui -sous certaines pressions environnementales- ont permis leur apparition.

A la recherche de ces capacités d'abstraction communes à l'homme et à l'animal, nous définirons ici les concepts de la manière globale et opérationnelle, pour pouvoir les étudier, comme: **la représentation mentale d'une catégorie**. Les catégories existent dans le monde qui nous entoure, les concepts sont des créations de notre esprit pour les traiter. Les niveaux d'abstraction sous-tendus, par les concepts animaux et humains et leurs différences, seront discutés au fur et à mesure de cette thèse.

1.5 Un monde de catégories

Le monde extérieur est d'une incroyable variété. Au cours d'une vie, chaque objet et situation nous apparaît selon des milliers de déclinaisons. Traiter indépendamment chacun de ces événements uniques est une tâche d'une complexité incroyable, hors de portée du plus puissant des ordinateurs actuels. Et pourtant, l'homme, et les autres animaux, s'adaptent et survivent dans ce monde en perpétuel changement. Comment? En détectant et en utilisant les invariants. Effectivement, rien n'est vraiment unique dans la nature, quelle que soit l'échelle que l'on considère: toute matière est composée d'atomes ; vivant et non vivant partagent certains éléments chimiques, animaux et végétaux, ont des cellules qui se ressemblent, la plupart des animaux se

déplacent, tous les oiseaux sont équipés d'un bec et d'un plumage et l'ensemble de l'univers est régi par des lois que la physique nous révèle petit à petit.
Au lieu de traiter chaque objet ou événement de manière unique, il est extrêmement économique d'en utiliser certains invariants, pour les regrouper en classes auxquelles on associe certaines caractéristiques et auxquelles nous pouvons réagir de la manière appropriée pour notre survie ou celle de notre espèce. C'est la capacité de catégorisation.

Bruner, Goodnow et Austin, dans un livre faisant référence en la matière, *A study of thinking* (1956), révèlent cinq propriétés de cette fonction cognitive :

1- Catégoriser permet, comme nous l'avons dit précédemment, de réduire la complexité de l'environnement. Il existe plusieurs millions de nuances de couleurs, si nous devions associer un nom à chacune d'elles, nous passerions notre vie à apprendre cette liste.
2- Catégoriser est le moyen par lequel nous identifions les objets. On reconnaît généralement une forme quand on arrive à la classer dans une catégorie familière telle que chien, voiture ou lettre Q.
3- Par conséquent, catégoriser permet aussi de réduire le besoin d'apprentissage continuel. On peut « comprendre » des nouveaux objets sans les « apprendre » si on arrive à les assimiler à une classe connue.
4- Catégoriser, c'est prendre des décisions appropriées. Nous n'aurons pas la même réaction si on classifie un objet sinueux comme un tuyau d'arrosage ou comme un serpent venimeux.
5- Enfin, catégoriser permet d'ordonner et de relier des classes d'objets entre elles : d'organiser notre connaissance.

Tâche d'identification de concept

Comment étudier cette capacité cognitive qui paraît si fondamentale pour structurer notre pensée? Les premières études sur la catégorisation chez l'homme en psychologie cognitive ont été menées dans des tâches dites d'identification de concepts (Bruner et al. 1956, Shepard et al. 1961). Le but de ce genre de tâche est d'évaluer si des sujets sont capables de regrouper spontanément des objets comme appartenant à un concept donné. Ces concepts étant généralement définis par des règles logiques (conjonctive, conditionnelle, etc.). Dans ce genre d'expériences les

concepts et les items sont artificiels. Les stimuli sont souvent des formes géométriques qui varient selon des dimensions évidentes et le concept à découvrir peut être, par exemple, d'identifier tous les stimuli rouges et de petite taille (règle conjonctive associant une couleur et une taille).

Ces tâches, volontairement épurées et abstraites, mises en place par les premiers psychologues cognitivistes pour étudier l'essence de la catégorisation, sont parfois loin des réalités et contraintes du monde naturel et leurs résultats sont difficiles à transférer au fonctionnement de l'esprit humain dans son environnement réel. C'est la principale critique de Rosch qui, dans les années 70, est à l'initiative de l'étude des catégories naturelles, qui correspondent à des objets ou événements existant dans le monde réel.

Hiérarchie des catégories naturelles

La principale contribution de Rosch est de remarquer que les catégories naturelles sont hiérarchisées: des catégories larges englobent plusieurs sous-catégories plus fines, comme des poupées russes. Un même objet du monde réel peut donc être catégorisé à différents niveaux. Elle en distingue trois :

- Les catégories supérieures ou superordonnées, les plus larges ; par exemple, les fruits, les instruments de musique, les véhicules ou les animaux.
- Les catégories de base, de niveau intermédiaire ; par exemple, les pommes, les guitares, les voitures ou les chiens.
- Les catégories subordonnées, les plus fines ; par exemple, les pommes golden, les guitares folk, les cabriolets ou les lévriers.

Niveau de catégorisation et degré d'abstraction

Dans son papier de référence publié en 1976, Rosch mène une série d'expériences (12 au total) pour montrer que le niveau basique est le niveau de catégorisation spontané, évident et courant pour l'être humain. Ces résultats sont mis en évidence en contrastant niveau de base et niveau superordonné. Ce dernier, de par la variabilité des exemplaires qui le composent, demande un effort d'abstraction supplémentaire pour être atteint, qui le place systématiquement à un niveau plus complexe que le niveau de base dans chacune des expériences. Nous ne détaillerons ici que trois

expériences particulièrement illustratives de cet avantage du niveau basique sur le niveau superordonné.

Dans la première expérience, les auteurs demandent à des sujets humains de quantifier le nombre d'attributs communs à chaque niveau pour différents types de catégories. L'analyse du nombre moyen d'attributs (Table 1) fait ressortir de manière très claire un parallèle entre niveau de hiérarchie et degré d'abstraction : le nombre d'attributs communs, pour chaque catégorie, s'accroît, depuis la catégorie supérieure jusqu'à la catégorie subordonnée. Catégoriser des objets au niveau supérieur demande d'ignorer beaucoup de détails de chaque objet pour ne se concentrer que sur les quelques attributs communs, c'est un niveau d'abstraction fort. Au contraire, au niveau de base et au niveau subordonné les objets présentent de nombreux attributs communs, la catégorisation à ces niveaux ne nécessitant qu'un faible niveau d'abstraction.

Catégorie	**Supérieur**	**Niveau de base**	**Subordonnée**
Fruits	3	8.3	9.5
Véhicule	1	11.7	16.8
Vêtement	2	8.3	9.7

Table 1: Nombre moyen d'attributs en commun à chaque niveau d'abstraction (tiré de Rosch et al 1976)

La table 2 donne un exemple du type d'attributs rapportés pour les vêtements. Notez que les attributs rapportés pour caractériser le niveau supérieur sont de type fonctionnel et nous renseignent peu sur l'aspect visuel de l'objet, alors que les attributs du niveau de base et subordonné sont principalement perceptifs.

Catégorie	**Supérieur**	**Niveau de base**	**Subordonnée**
Vêtement	**Vêtement** - Porté habituellement - Pour se protéger du froid	**Pantalons** - Pour les jambes - Les boutons - Les passants - Les poches - Le tissu - Deux jambes	**Jeans** - Bleu **Pantalon en toile** - Confortable - Extensible

Table 2: Exemples d'attributs partagés à chaque niveau pour les vêtements (tiré de Rosch et al 1976). Seul les attributs supplémentaires par rapport aux autres niveaux sont rapportés.

La théorie prévalant à l'époque pour expliquer le fonctionnement de la catégorisation est celle du prototype. Pour catégoriser de nouvelles formes, on les compare à des prototypes que nous avons appris. Ces prototypes sont des représentations mentales du stimulus « moyen » correspondant à chaque catégorie. En se basant sur cette définition, Rosch et al. montrent, dans la deuxième expérience, que les sujets n'arrivent pas bien à identifier la forme moyenne entre différents exemplaires du niveau supérieur (moyenne entre une chemise, un pantalon et des chaussettes) alors qu'ils sont parfaitement capables de le faire au niveau inferieur (moyenne entre différents types de pantalons). On peut donc identifier un prototype moyen au niveau de base et subordonné, mais pas au niveau supérieur.

Enfin dans la troisième expérience dont nous avons choisi de parler ici, les auteurs s'intéressent à l'acquisition des niveaux de catégorisation au cours de l'enfance. Pour ce faire, ils proposent à des enfants de 3 à 10 ans de classer des triades de photographies couleurs en montrant du doigt celles qui se ressemblent. En fonction des images qu'ils présentent, Rosch et al. peuvent évaluer leur capacité à former des catégories au niveau basique ou au niveau superordonné. Les résultats sont clairs et nets. Alors que dès 3 ans, les enfants sont capables de trier correctement, avec un pourcentage correct de 99%, les images au niveau basique, au même âge ils ne classifient correctement les images au niveau superordonné que dans 55% des cas. Ce n'est qu'à 4 ans que les enfants atteignent 96% de bonnes réponses au niveau superordonné. Les auteurs démontrent que cette différence n'est pas due à l'acquisition du langage.

L'ensemble des résultats de Rosch démontre, comme elle le défend, qu'il existe chez l'homme un traitement spontané et privilégié des objets naturels au niveau basique. Tout au long de cette démonstration on s'aperçoit que classer les objets au niveau supérieur, est en revanche une tâche plus difficile, qui apparait plus tard dans le développement de l'homme et nécessite une plus grande capacité d'abstraction visuelle.

Toutefois, des études récentes (Macé et al. 2009, Joubert et al. 2007) semblent remettre en cause cet avantage de la catégorie de base lorsque l'on considère uniquement des représentations visuelles. D'après ces auteurs (Fabre-Thorpe 2011), le langage jouerait un rôle déterminant dans la latence d'accès à la catégorie de base.

Catégories perceptuelles ou catégories basées sur des règles

Une autre manière de distinguer les types de catégories, consiste à opposer les catégories basées sur des similarités physiques ou perceptuelles (BSP) aux catégories indépendantes des similarités physiques (ISP). Les catégories BSP englobent toutes les catégories dont les exemplaires partagent au moins un trait perceptuel (couleur, texture, forme, luminance, tonalité, odeur, goût, etc.). Les catégories subordonnées et basiques appartiennent clairement aux catégories BSP.

Au contraire, les membres d'une catégorie ISP ne partagent, par définition, pas de trait physique, c'est le cas des catégories multimodales correspondant par exemple au concept d'une personne : sa voix, son visage, son nom. Les catégories dites basées sur des règles abstraites (BRA) peuvent aussi entrer dans la classe des catégories ISP. Effectivement, les catégories BRA n'impliquent aucune similarité physique entre ses membres, puisque les liens catégoriels reposent sur des critères plus abstraits de type fonctionnel (objets comestibles) ou relationnel (dessus/dessous).

Ces dernières, les catégories BRA, s'opposent aussi aux catégories BSP par la manière dont elles sont apprises. On admet qu'une catégorie BRA nécessite un apprentissage supervisé ; lorsqu'un individu classifie un stimulus, un processus de rétrocontrôle le renseigne sur la validité de son choix. Ce processus fait découvrir à l'individu la règle imposée par l'expérimentateur lui permettant de catégoriser correctement les stimuli. Au contraire, on attribue l'acquisition de catégories BSP à un apprentissage non supervisé, qui permet à l'individu de construire une catégorie par lui-même, sans l'aide d'un expérimentateur ou d'un professeur. Les membres d'une catégorie BSP, de par leurs similarités conjointes et leurs différences par rapport à d'autres stimuli, peuvent être regroupés spontanément sans rétrocontrôle.

Il est assez tentant pour simplifier l'étude de la catégorisation notamment chez l'animal, d'opposer systématiquement catégories BSP et catégories BRA. Beaucoup

de chercheurs ont hiérarchisé ces deux types de catégories en attribuant une faible demande en abstraction pour former des catégories BSP, alors que la catégorie BRA, elle, témoignerait de capacités d'abstraction supérieures. La distinction n'est pas si facile. Nous avons vu qu'au sein des catégories naturelles, qui sont généralement basées sur des similarités, il existait plusieurs niveaux d'abstraction, notamment le niveau supérieur qui présente un niveau d'abstraction tel qu'il ne peut parfois être décrit verbalement que par des relations de type fonctionnel: les outils sont des objets qui servent à construire, pourtant ils partagent aussi des similarités physiques comme un manche, le bois, l'acier ou le plastique. De plus, les deux mécanismes d'apprentissage de ces deux types de catégories peuvent être mis en jeu à différents moments de l'acquisition d'une catégorie. Comme l'a montré Behl-Chadah (1995), les enfants regroupent très tôt de manière spontanée les mammifères terrestres ensemble, mais c'est seulement parce que nos parents ou nos professeurs nous l'ont dit que l'on catégorise les baleines avec les mammifères et non avec les poissons. Ainsi, selon les auteurs, les catégories naturelles supérieures appartiendront soit aux catégories BSP soit aux catégories BRA.

2 Cognition comparée de l'abstraction visuelle :

2.1 Introduction

2.1.1 Les 4 questions de Tinbergen : un cadre théorique pour étudier le comportement animal

L'étude du comportement animal a de tout temps passionné biologistes et naturalistes. Mais ce n'est que dans les années 50 que l'éthologie lui donne ses lettres de noblesse. Konrad Lorenz et Nikolaas Tinbergen sont les premiers à fournir un cadre conceptuel suffisamment précis et complet pour comprendre et interpréter les comportements animaux dans leurs subtilités et leurs richesses. On doit notamment à Nikolaas Tinbergen ses 4 questions permettant l'étude approfondie des comportements. Pour bien comprendre ces questions, la table 3 propose de les présenter comme la combinaison de deux facteurs : celui des *causes du comportement* et celui de *l'échelle temporelle*.

		Perspective temporelle: Diachronique vs Synchronique	
		Vue dynamique Expliquer le comportement par son historique	**Vue Statique** Expliquer le comportement selon des facteurs actuels
Causes: Comment vs Pourquoi?	**Cause proximale** **Comment** fonctionne le comportement d'un individu?	**Ontogénie** Comment le comportement se crée au cours du développement de **l'individu** ?	**Mécanisme** Quels sont les mécanismes cognitifs et neuronaux permettant le comportement ?
	Cause distale (évolutionniste) **Pourquoi** une espèce possède ce comportement?	**Phylogénie** Quelle est l'histoire évolutive qui a amené l'**espèce** à posséder ce comportement?	**Fonction (adaptation)** A quoi sert le comportement en termes adaptatifs dans l'environnement actuel de l'animal? Quelle est sa valeur de survie ?

Table 3: Les quatre questions de Nikolaas Tinbergen pour l'étude et l'interprétation des comportements

Pour ce qui est des causes, les *causes proximales* sont opposées aux *causes ultimes*.

L'étude des causes proximales est synonyme d'étude mécanistique du comportement : on s'intéresse aux facteurs biologiques et cognitifs qui permettent à un individu d'exprimer ce comportement. Cette approche proximale peut être abordée de manière

statique, sur l'individu tel qu'il est à l'instant présent, c'est l'objet des neurosciences (cognitives) et de la psychologie cognitive. Dans le cadre de cette thèse, l'approche mécanistique statique consiste en l'étude des capacités cognitives (perception visuelle, reconnaissance, catégorisation, décision) sous-jacentes à la catégorisation rapide ainsi que celle des réseaux et mécanismes neuronaux qui la sous-tendent.

L'approche *proximale dynamique* est l'étude développementale du comportement, elle appartient à la biologie et à la psychologie du développement. On s'intéresse à tous les événements, depuis l'embryogénèse jusqu'au vieillissement, qui contribuent à l'évolution du comportement. Par exemple, l'étude du développement de la reconnaissance visuelle nous enseigne que les nourrissons peuvent, dès 6 mois, distinguer des visages d'individus d'une autre espèce mais perdent cette capacité lorsqu'ils atteignent 9 mois (Pascalis et al 2002), qu'au bout de deux ans il peuvent former des catégories basiques et que c'est entre 3 et 4 ans qu'ils accèdent au niveau de catégorisation superordonnée (Rosch et al 1976).

A l'étude des causes proximales, s'oppose l'étude des *causes ultimes*, *distales* ou *évolutives* du comportement. Au lieu de chercher les mécanismes biologiques internes sous-tendant le comportement, l'approche évolutive cherche à élucider les mécanismes environnementaux qui en sont à l'origine.

Là encore on peut adopter une approche *statique*: à quoi sert le comportement du sujet dans son environnement actuel, quelle est sa fonction? Pour nos sujets humains, catégoriser des animaux rapidement leur permet de faire de l'observation naturaliste ou bien, au volant d'un véhicule, de décider si un objet au bord de la route est un animal, et présente donc un risque de traverser devant eux au dernier moment, ou si ce n'est qu'un buisson. Pour nos singes, catégoriser des images comme contenant des animaux, est tout simplement un moyen d'obtenir une récompense hydratante et délicieusement sucrée. Dans les deux exemples cités précédemment, on se rend bien compte que les fonctions ne sont pas des plus « adaptatives » ou naturelles qu'il soit.

Pour obtenir une raison plus convaincante de l'existence de la capacité de catégorisation rapide, il faudra donc se tourner vers une approche *dynamique* des causes distales : c'est l'étude phylogénétique du comportement. Quels sont les

facteurs environnementaux qui ont façonné la capacité des primates à reconnaître des animaux en une fraction de seconde ? Il est crucial, pour la survie d'un primate et celle de son espèce, de catégoriser le plus rapidement possible un objet inconnu comme animal ou non animal, prédateur ou congénère.

Dans cette thèse nous essaierons de discuter des 4 niveaux d'analyse du comportement proposés par Tinbergen ; néanmoins, les expériences menées dans le cadre du laboratoire sur les singes et sujets humains adultes ne peuvent documenter que l'approche mécanistique statique du comportement.

2.1.2 La cognition comparée : définition, enjeux et limites

La cognition comparée est une discipline dont les objectifs précis et les limites sont difficiles à appréhender. Comme nous l'avons vu dans notre historique, c'est un mélange contemporain de l'éthologie cognitive et de la psychologie comparée. Néanmoins, une définition rudimentaire pourrait être : l'étude phylogénétique de la pensée.

Les espèces animales ont d'abord été décrites au niveau macroscopique en caractérisant leurs phénotypes, c'est-à-dire en définissant leurs caractères morphologiques, anatomiques et physiologiques. Avec la découverte de l'ADN on peut maintenant décrire une espèce au niveau moléculaire, grâce à son génotype. De la même manière, l'éthologie classique, en recensant l'ensemble des comportements mis en place dans la nature (éthogramme), a tenté de définir l'espèce au niveau comportemental.

Mais, qu'il s'agisse de gènes, d'organes ou de comportements, le biologiste, malgré son désir ardent de classifier le vivant, ne peut que constater que nombre de ces marqueurs sont partagés par plusieurs espèces. Ce constat, qui a amené Darwin à proposer sa théorie de l'évolution, a donné naissance à l'étude des phylogénies. Cette discipline, en comparant les organes ou gènes partagés par les espèces, vise à établir les liens de parenté qui les lient et ainsi à retracer l'histoire du vivant.

De la même manière, la cognition comparée défend que la variété des comportements observés à travers les différentes espèces, peut en fait s'expliquer par l'utilisation d'un jeu -plus ou moins complet- de fonctions cognitives -plus moins développées- commun entre les espèces. Ce matériel cognitif commun devrait s'ajouter à la liste des marqueurs du biologiste pour l'étude des phylogénies.

Concrètement, le passe-temps favori des chercheurs en cognition comparée, est la démonstration de capacités animales alors qu'elles étaient dites, jusqu'alors, spécifiquement « humaines ». Un exemple frappant est celui de l'utilisation d'outils. Il y a encore une quarantaine d'années, personne (dans les sphères scientifiques et philosophiques, du moins) n'aurait imaginé que les animaux soient capables d'utiliser des outils. C'était certainement une des capacités majeures permettant de nous différencier du reste du vivant. C'est grâce à l'outil que l'homme a pu petit à petit s'abstraire des aléas de la nature et fonder sa civilisation: l'outil pour se protéger, l'outil pour chasser, l'outil pour cultiver, l'outil pour construire. Or, dans les années 60, à l'instigation du paléoanthropologue Louis Leakey, une jeune étudiante part en Tanzanie, à Gombe, pour y observer un groupe de chimpanzés sauvages. Au cours de ces nombreuses années d'étude de terrain, Jane Goodall rapportera un ensemble de données révolutionnaires qui ont changé à tout jamais notre façon de voir les animaux. Parmi elles, elle relate l'utilisation naturelle de brindilles que les singes choisissent minutieusement et qu'ils effeuillent si besoin, pour « pêcher » les termites. Les animaux sont donc capables d'utiliser des outils. Ce résultat extraordinaire en engendrera bien d'autres ; on montrera notamment que les chimpanzés utilisent aussi des kit-outils (une enclume et un marteau) pour casser des noix, qu'ils se servent parfois de projectiles pour chasser ; mais aussi que beaucoup d'autres espèces sont capables d'utiliser des outils, comme les capucins moines, certains singes du nouveau monde, les loutres de mer, certaines espèces de pinsons, les grands dauphins, plusieurs espèces de corvidés, l'ours brun et le dingo d'Australie (pour ne citer que les exemples que j'ai en tête). En fait, il se pourrait bien que la plupart des animaux soient capables d'utiliser des outils dans certaines situations particulières. Ces résultats, en ancrant l'homme, dans l'évolution animale, par ses capacités cognitives les plus « humaines », sont des arguments forts pour l'hypothèse de continuité.

Même si Darwin postule, dès le XIXe siècle une continuité des fonctions de l'esprit dans le règne animal, la cognition comparée reste une discipline très récente qui a bien évidemment dû attendre le « boum » du cognitivisme dans la deuxième partie du XXe siècle pour commencer à émerger. Et ce n'est que depuis une trentaine d'années, que cette approche s'est construite sur des bases méthodologiques solides. Il n'est donc pas étonnant que cette jeune discipline souffre encore de certaines difficultés théoriques.

La première, inhérente à notre nature humaine, est l'anthropocentrisme. En effet, beaucoup de chercheurs s'intéressent à la cognition animale, uniquement pour expliquer les origines de la cognition humaine. Il s'agit là d'un parti pris dans le choix du sujet d'études (qui, en passant, n'est certainement pas indépendant de la capacité de ce sujet à générer des publications à haut *impact factor*). Néanmoins, cette approche peut aussi être un handicap et un facteur limitant pour aborder de manière objective l'étude de la cognition animale (homme compris) et de son évolution. Il est évident qu'à l'instar de la variété morphologique observable dans le règne animal, l'évolution a dû façonner une variété de fonctions cognitives que nous ne possédons pas. L'exemple le plus marquant est certainement celui de la capacité d'écholocalisation présente chez les chauves-souris et certains mammifères marins. C'est une capacité sensorielle que nous ne possédons pas (excepté pour de rares cas particuliers de personnes aveugles qui utilisent l'écho de leurs pas pour localiser la distance entre eux et des objets), et qui permet de localiser des objets au moyen d'émissions et de réceptions d'ondes acoustiques. Ce type de capacité sensorielle est extrêmement difficile à imaginer de par notre absence d'expérience subjective en la matière. Ce n'est que par l'observation objective de chauves-souris dans leur milieu naturel, suivie par des expérimentations rigoureuses, que Lazzaro Spallanzani, puis beaucoup plus tard Donald Griffin, ont pu mettre en évidence cette capacité typiquement non humaine. Il est extrêmement difficile de se sortir du carcan de notre subjectivité humaine pour découvrir d'autres mondes cognitifs, et pourtant c'est un exercice qu'à mon avis le cognitiviste comparatiste se doit d'effectuer autant que possible.

Une autre difficulté pour la cognition comparée est le faible nombre d'espèces étudiées. Si on la compare avec l'approche biologique (anatomique) comparée

classique, la cognition comparée en est à peine au stade de "classification". Dans cette première étape nécessaire pour évaluer le potentiel cognitif du règne animal dans son ensemble, il reste encore un travail considérable à effectuer. Effectivement, comme nous le verrons dans les chapitres suivants, seules quelques espèces animales ont fait l'objet d'études poussées visant à étudier un bon nombre de leurs capacités cognitives; c'est notamment le cas de l'abeille, de la drosophile, du pigeon, du perroquet gris du Gabon, du babouin olive, du macaque rhésus, du capucin moine, du chimpanzé et du grand dauphin … La plupart des autres espèces étudiées le sont pour une capacité cognitive particulière qui les caractérise. On étudie les geais buissonniers pour leur aptitude à faire des réserves de nourriture révélant leur remarquable mémoire, et leur capacité à voyager dans le temps mentalement, ou encore les corneilles de Nouvelle-Calédonie pour leur capacité à façonner et utiliser des outils à base de végétaux, dans le but de capturer des larves d'insectes. Ces quelques espèces ne représentent qu'une infime proportion de l'ensemble des espèces connues.

Même si son travail de classification est loin d'être terminé, la cognition comparée se veut, comme son nom l'indique, "comparative", et c'est certainement la partie la plus délicate de cette science. D'abord parce que s'il est possible de décrire entièrement les espèces par des génotypes ou par des phénotypes exhaustifs, établir des « cognotypes » complets est une tâche titanesque. Les éthologues proposent pour certaines espèces des éthogrammes qui recensent l'ensemble des comportements présentés dans la nature par une espèce. Ce genre de recensement est déjà le fruit d'années de travail et d'observations minutieuses, mais établir toutes les fonctions cognitives, et leurs degrés de développement sous-tendant ces comportements, nécessite des dizaines d'années supplémentaires d'expérimentations et d'observations. Or, pour pouvoir retracer l'histoire détaillée de la cognition il serait nécessaire de posséder l'inventaire le plus complet possible des capacités cognitives de chaque espèce.

Enfin, la cognition est la fonction, le cerveau est le substrat biologique. Etudier et utiliser la cognition seule pour déduire l'existence d'une proximité phylogénétique entre deux espèces revient à dire que la guêpe et le rouge-gorge sont de proches parents juste parce qu'ils volent tous les deux. Une analyse plus détaillée de l'organe qui permet à ces deux espèces de voler, montrera que si l'aile de la guêpe et celle du

rouge-gorge remplissent une fonction analogue, c'est probablement dû à des évolutions indépendantes, sous des pressions du milieu similaires, plutôt qu'à l'existence d'un caractère homologue entre ces deux espèces, révélant un ancêtre commun. Au contraire, la nageoire du dauphin, la patte avant du chien et le bras de l'homme remplissent des fonctions très différentes. Pourtant, une analyse minutieuse des os qui les constituent, révèle une origine commune mais des adaptations morphologiques différentes en fonction des pressions du milieu différente dans lesquelles chacune de ces espèces a évolué. Pour moi, la cognition comparée doit donc s'allier aux neurosciences comparées pour permettre d'établir l'arbre évolutif de la vie mentale des animaux. C'est une des raisons pour lesquelles l'approche comportementale développée dans cette première partie est doublée d'une approche neuronale présentée dans la troisième partie.

2.1.3 Les méthodes pour étudier l'abstraction chez l'animal

Cognition spontanée ou apprentissage ?

Le principal écueil auquel se heurte celui qui veut étudier les capacités mentales des animaux est l'absence de langage commun entre eux et nous. Il y a deux possibilités pour pallier ce handicap:

- Soit observer les animaux dans leur milieu naturel, analyser leurs comportements et essayer d'en déduire quelles fonctions cognitives ils sous-tendent. C'est l'approche classique de l'éthologie qu'on appelle aussi maintenant écologie comportementale. Cette approche permet de mettre en évidence des capacités cognitives « naturelles » et elle est idéale pour comprendre les causes distales des comportements ainsi que leurs fonctions évolutive et écologique.
- Néanmoins, on a bien souvent envie d'aller plus loin et de réellement 'interroger' l'animal dans un cadre expérimental, dans lequel on contrôle un maximum de paramètres. Si cette approche est clairement celle des behaviouristes des premières heures, elle est restée ancrée dans les méthodes de la cognition comparée moderne. C'est dans ce cadre expérimental que tous les résultats de cette thèse ont été acquis.

Selon la stratégie expérimentale mise en place, différents aspects de la cognition sont évalués. On peut, par exemple, s'intéresser aux catégories que l'animal forme spontanément ou à celles qu'il est capable d'apprendre.

Dans le premier cas, on étudie le comportement par défaut, ou naturel, des animaux. Dans le contexte de la cognition visuelle qui nous intéresse ici, la variable la plus souvent étudiée est le temps de regard (préférentiel), une approche initialement développée chez les nouveau-nés (Teller 1979). On présente un ou deux stimuli au sujet et on enregistre le temps qu'il a passé à le(s) regarder. S'il passe significativement plus de temps à regarder un type de stimulus plutôt qu'un autre, on en déduira qu'il est capable de dissocier ces deux types de stimuli.

Dans un exemple de protocole utilisant la présentation d'une seule image, une équipe allemande, travaillant sur les macaques de barbarie de Rocamadour, a montré que les adultes reconnaissent spontanément les individus de leur propre groupe (Schell et al. 2011). En effet, ils observent significativement moins longtemps la photo d'un membre de leur groupe que celle d'un macaque d'un autre groupe. Ce n'est pas le cas des macaques juvéniles.

Néanmoins, dans ce genre d'approche, les animaux sont tellement libres qu'il est parfois difficile d'inférer les mécanismes cognitifs sous-jacents. Une version un peu plus raffinée de ce protocole, appelée comparaison visuelle appariée, consiste à habituer un sujet à un type de stimulus, puis à lui présenter une nouvelle paire de stimuli, avec un stimulus familier et un autre stimulus. Pascalis et al. (2002) montrent ainsi que les enfants de 9 mois et les adultes regardent plus longtemps les visages humains nouveaux mais ne montrent pas de préférence pour des visages de singes. Par contre, les enfants de 6 mois préfèrent regarder des visages nouveaux qu'ils soient de singes ou d'hommes. Ces résultats dénotent une capacité à discriminer des individus d'une autre espèce, présente chez le nouveau-né mais qui disparaît en grandissant.

Le temps de regard est aussi utilisé pour montrer des capacités à former spontanément des concepts cross-modaux. Dans ce cas, on analyse l'effet de la présentation d'un stimulus non visuel (auditif par exemple) sur les temps de regard de stimuli visuels.

C'est ainsi que Sliwa et al. (2011) révèlent que les macaques forment spontanément un concept d'identité pour des congénères, mais aussi pour des humains familiers. Confrontés à deux visages, dont l'un correspond au cri (la voix) qu'ils viennent d'entendre, ils passent plus de temps à observer le visage congruent avec ce cri.

A l'opposé, un protocole de violation de l'attente, permet d'évaluer sur le terrain l'existence de concepts cross-modaux pour des animaux qu'il est difficile de faire asseoir derrière un écran d'ordinateur. Par exemple, si l'on fait passer devant un cheval un de ses congénères et que, dès que ce dernier disparaît de son champ de vision, un hennissement, qui ne correspond pas au cheval qui vient de passer, est émis par un haut parleur ; l'animal testé va regarder plus rapidement et plus longtemps dans la direction de la disparition de son congénère, que si le hennissement est congruent avec l'individu qui vient de passer (Proops et al 2009).

Ces protocoles de temps de regard permettent d'étudier les concepts naturels. Ils sont rapides à mettre en place puisqu'ils ne demandent pas (ou peu) de procédure d'apprentissage ni d'entraînement de l'animal. Ils ne nécessitent qu'une logistique modérée facilement adaptable aux conditions de vie des différentes espèces ; et surtout, ils peuvent être utilisés sans contention et privation alimentaire ou hydrique de l'animal. Cet outil expérimental, encore relativement peu utilisé pour l'étude la catégorisation, est sans nul doute amené à devenir un des fers de lance de la cognition comparée permettant de tester un grand nombre d'espèces en très peu de temps.

Néanmoins, pour des raisons historiques, et notamment sous l'influence du behaviourisme, la plupart des recherches de ces 50 dernières années en catégorisation chez l'animal ont été effectuées en utilisant le protocole classique de conditionnement opérant développé par Skinner au début du XXème siècle. L'animal est maintenu dans un dispositif de test qui comporte: un système d'affichage (écran d'ordinateur ou projecteur de diapositives, pour les premières expériences des années 60/70), un système permettant d'enregistrer la réponse catégorielle et un distributeur de récompense (nourriture ou boisson). Ainsi placé dans le dispositif, l'animal devra fournir une réponse particulière associée à une catégorie d'images présentées. Chez le pigeon, une des premières espèces dont les capacités de catégorisation ont été étudiées, la réponse est généralement la fréquence de picorage d'une clé de réponse,

selon la catégorie d'image présentée (Herrnstein et Loveland 1964) ou bien le picorage direct d'un stimulus particulier parmi plusieurs autres possibles (Cook et al. 2000). Pour les singes, les réponses demandées sont variables. Ils peuvent être amenés à toucher un carré de réponse sur un écran tactile (Matsuzawa 1985), à relâcher un bouton et toucher le stimulus (Fabre-Thorpe et al. 1998), à utiliser un joystick (Fagot et al. 2001) ou encore à produire des saccades oculaires vers un type de stimulus donné (Vogels 1999, Girard et al. 2008). Pour réaliser ces tâches, les animaux sont motivés par une récompense associée à chaque bonne réponse. Contrairement aux observations de terrain ou aux temps de regard, il faut être clair avec le fait que les protocoles de conditionnement opérant, testent la capacité d'un animal à apprendre une catégorie prédéfinie par l'expérimentateur et non une catégorie qu'il formerait spontanément dans son milieu naturel. C'est la principale limitation de cette technique pour étudier la cognition animale en tant que telle, mais c'est aussi son point fort pour permettre la comparaison entre espèces. En effet, comme dans l'étude présentée ici, on peut évaluer, par exemple, à quel point les singes sont capables d'apprendre des catégories qui ont un sens pour l'homme.

Il est intéressant de noter que, certainement sous la pression sociétale qui ne porte pas très haut dans son estime l'expérimentation animale, une nouvelle manière de faire de la psychologie animale expérimentale est en train de voir le jour. A côté d'Aix en Provence, dans une station de primatologie appartenant au CNRS, des chercheurs bricoleurs ont mis au point un système permettant de tester les capacités cognitives d'un groupe de 26 babouins en semi-liberté (Figure 1). Ils utilisent des protocoles de psychologie comparée classiques, présentés sur écran d'ordinateur, mais sans contraindre les animaux et sans les priver de quoi que ce soit (Fagot et Bonté 2010). Leur secret : ils ont construit 10 postes de travail attenants à l'enclos du groupe où tous les individus peuvent venir gagner du blé, en guise de récompense. Les animaux sont donc en permanence au sein de leur groupe social et ils travaillent quand et autant qu'ils le souhaitent, en ayant une alimentation et un régime hydrique tout à fait normal. Grâce à une petite puce électronique placée chirurgicalement sous la peau, chaque individu peut être identifié lorsqu'il entre dans le dispositif expérimental et l'ordinateur peut proposer directement le protocole choisi pour cet individu en particulier. Si ce dispositif constitue le nec plus ultra de l'expérimentation animale en matière d'éthique et de bien-être animal, il fournit aussi des données scientifiques

sans précédent. Ainsi, il permet de tester, pour la première fois en psychologie animale expérimentale, une véritable population d'individus de 2 à 32 ans, qui travaillent presque en continu (Figure 1d), et d'obtenir un très grand nombre d'essais (plus d'1 million en 85 jours d'expérience). Il est indéniable que ce genre de dispositif constitue l'outil du futur pour l'étude de la cognition animale nécessitant un conditionnement opérant et qu'il est amené à produire rapidement une recherche quantitative au pouvoir statistique fort et de grande qualité scientifique, à l'image du papier récemment publié dans Science par le groupe, montrant la capacité des babouins à distinguer des mots de non-mots (Grainger et al 2012).

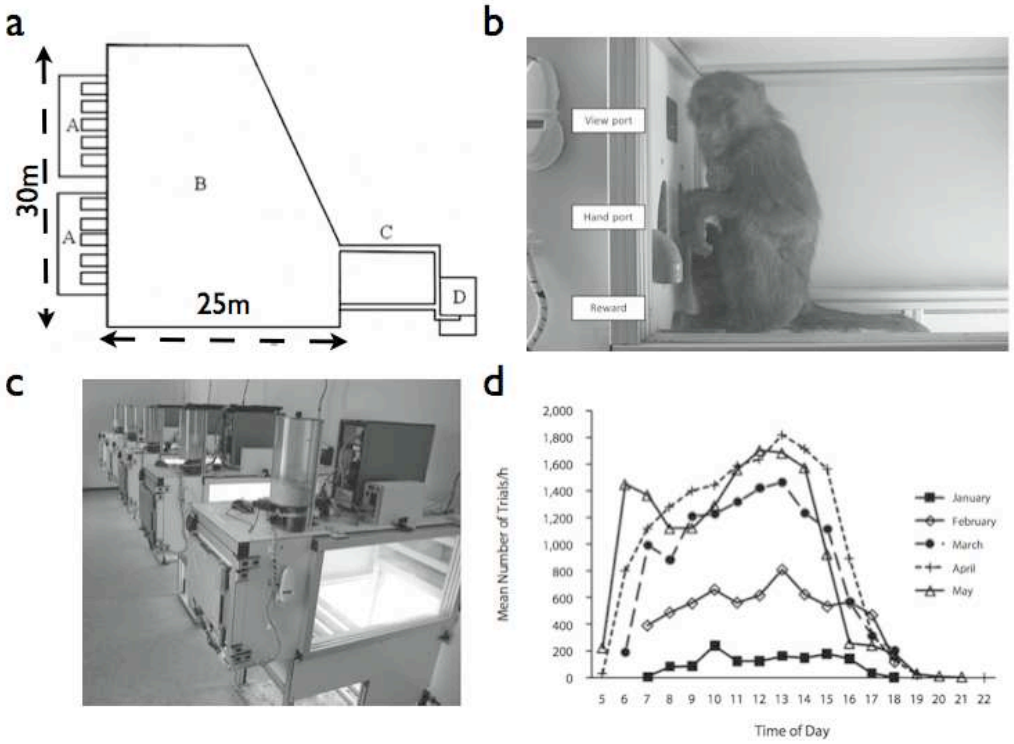

Figure 1: Dispositif d'expérimentation automatique pour un groupe de 26 babouins en semi-liberté (adapté de Fagot et Bonté 2010)

a) Dix postes d'expérimentation (A) sont mis à disposition du groupe hébergé dans un parc en plein air (B) où des tunnels (C) leur permettent de rejoindre un abri (D).
b) Jeune babouin au travail dans un des postes. Sont indiqués le système de délivrance de récompense (reward), les trous permettant de voir l'écran (view port) et de passer les mains (hand port) pour répondre.
c) 5 des postes de travail.
d) Evolution du nombre d'essais (effectués par l'ensemble des individus) en fonction de l'heure du jour (abscisse) et du mois (différentes courbes).

Enfin, on peut aussi utiliser le cadre expérimental et le conditionnement opérant pour essayer de tester les catégorisations spontanées des animaux. Sands et al (1982) ont entraîné des macaques rhésus à réaliser une tâche où ils devaient répondre quand deux images successives leur paraissaient identiques. Il n'y avait donc pas de consigne explicite sur le type de catégorie à former. Les stimuli étaient des photos de visages humains, de visages de macaques, d'arbres, de fruits et de fleurs. En analysant les patterns d'erreur, les chercheurs ont clairement montré que les membres d'une même catégorie étaient perçus par les singes comme plus similaires entre eux qu'avec les membres d'autres catégories. Chez le pigeon, Wasserman et al (1988) compare le temps d'apprentissage de catégories basiques basées sur des similarités physiques (voiture, chaise, chats et fleurs) avec celui associé à des pseudo-catégories composées de manière arbitraire d'un nombre équivalent d'exemplaires appartenant à ces 4 catégories. En 10 jours, les pigeons atteignent un score de 40% de bonnes réponses pour les pseudo-catégories contre 80% pour les catégories réelles. Les catégories humaines basiques semblent donc aussi se structurer de façon préférentielle chez les pigeons.

Discrimination, généralisation et catégorisation

Il faut bien distinguer les notions de discrimination et de catégorisation. Un organisme est capable de discriminer deux stimuli a et b s'il produit une réponse comportementale différente pour chacun de ces stimuli *en particulier*. Cette approche est généralement utilisée pour tester la capacité des animaux à percevoir certaines dimensions physiques de l'environnement (contraste, longueur d'onde…), ainsi que pour évaluer leur seuil de perception pour ces dimensions. Par exemple, Giurfa et al. (1997) entraînent des abeilles à associer une couleur (longueur d'onde) avec une solution sucrée. Dans la phase de test, ils évaluent leurs capacités à discriminer cette couleur en leur présentant la couleur sur laquelle elles ont été entraînées et une autre couleur, dans deux bras d'un labyrinthe en Y. La discrimination est donc une association *exemplaire*-réponse comportementale alors que la catégorisation est une association *classe*-réponse comportementale.

On dira qu'un animal *catégorise* les stimuli d'une *classe* A contre une classe B, s'il est capable de produire les mêmes réponses pour tous les exemplaires de la classe A et des réponses différentes pour tous les exemplaires de la classe B. Une notion clé pour montrer l'existence réelle de capacités de catégorisation est celle de *généralisation*. La plupart des expériences de catégorisation chez l'animal se déroulent en deux phases : une phase d'apprentissage où l'animal est entraîné à dissocier les classes de stimuli sur un set qui devient familier et une phase de test où l'on évalue sa capacité à généraliser en lui présentant des stimuli appartenant aux classes d'entraînement mais qu'il n'a jamais vus auparavant. Cette capacité de transfert à des stimuli nouveaux est nécessaire pour éliminer l'hypothèse d'apprentissage par coeur des associations stimulus-réponse catégorielles ayant été établies sur le set familier. Effectivement, il a été montré que les pigeons et les babouins par exemple, étaient capables de retenir respectivement plusieurs centaines et plusieurs milliers d'associations stimulus-réponse (Fagot et Cook 2009) même avec un seul essai en guise d'entraînement (Cook et Fagot 2009). Notons que nous utilisons ici le terme de généralisation au sens strict de transfert à de nouveaux stimuli, et non comme il a pu être précédemment utilisé (Benard et al 2006) pour qualifier un processus continu ou graduel, face au processus discret de la catégorisation.

Enfin, le dernier prérequis pour montrer avec certitude la catégorisation, corollaire de la notion du test de généralisation, est la discrimination entre stimuli d'entraînement et de test. Peu d'études évaluent vraiment cette capacité de discrimination intra-classe, car cela requiert un apprentissage supplémentaire difficile puisqu'il est orthogonal à la tâche initiale de catégorisation. Au moins une étude chez le babouin s'est attachée à remplir ces deux critères de généralisation et de discrimination intra-classe (Vauclair et Fagot 1996). Ils ont d'abord appris à des babouins à catégoriser des chiffres '3' contre des lettres 'B'. Dans un premier temps, en les testant avec une police d'écriture nouvelle pour les singes, ils ont montré leur capacité à généraliser. Dans un second temps, en utilisant une tâche d'appariement où le singe devait distinguer deux typographies de la même lettre B ou chiffre 3, il ont confirmé que les singes étaient bien capables de faire une tâche de discrimination intra-classe.

2.1.4 L'abstraction visuelle: une fonction largement répandue au sein du règne animal

Le pigeon: modèle de référence

L'histoire des relations entre les pigeons et l'homme date de la Perse ancienne. Selon Frontinus, leur capacité à transmettre une information rapidement à travers une longue distance aurait joué un rôle important dans la conquête de la Gaule par Jules Cesar. On dit aussi que les victoires des athlètes aux Jeux Olympiques antiques étaient transmises à leurs villes respectives par pigeons voyageurs. Ce n'est donc pas par hasard si cet ancêtre du téléphone, aux capacités d'orientation spatiale remarquables, a été, et est encore un des animaux les plus étudiés en cognition. En ce qui concerne l'étude de la catégorisation, on peut même dire que c'est par les pigeons que tout a commencé.

En 1964, l'expérience novatrice d'Herrnstein et Loveland, montrant, qu'après apprentissage, les pigeons pouvaient catégoriser des images nouvelles comme contenant ou non un être humain, posait la première brique de l'étude de l'abstraction animale. Depuis, des centaines de pigeons principalement dans des laboratoires américains mais aussi autrichiens ont été les sujets de dizaines d'expériences afin de décrire plus en détail leur capacité à catégoriser le monde qui les entoure. Dans la lignée des premiers travaux sur les catégories naturelles, il a ainsi été montré que les pigeons étaient capables de moduler leur fréquence de picorage pour différents types de catégories naturelles basées sur des similarités physiques ou basiques selon la terminologie de Rosch, comme les arbres, les poissons ou encore des conspécifiques (Aust & Huber 2002 ; Ghosh et al. 2004). Ils ont aussi réalisé avec succès des catégorisations naturelles de niveau superordonné comme : présence/absence d'eau (Herrnstein et al. 1976), animal/non animal (Roberts et Mazmanian 1988), artificiel/naturel (Lazareva et al. 2004) ou encore absence/présence de nourriture (Awazu et Fujita 2006).

A la remarque de Premack (1976), qui affirmait que seuls les primates pouvaient 'trier le monde', c'est-à-dire former des catégorisations multiples, et que les autres animaux n'étaient capables que d'une catégorisation de type présence/absence, les

spécialistes des pigeons ont répondu en développant une nouvelle tâche de catégorisation à choix forcés multiples. Dans ce protocole, le pigeon est entraîné à classifier plusieurs catégories non complémentaires (par exemple des chats, des fleurs, des chaises et des voitures, Bhatt et al. 1988) en les associant avec le picorage de 4 clés de réponses différentes. Ce protocole a été récemment poussé à l'extrême en montrant que les pigeons étaient capables de faire jusqu'à 16 catégorisations au niveau basique (bébé, trousseau de clé, canard, téléphone, etc.) dans la même expérience (Wasserman et al. 2007).

Encore plus intéressant, une étude récente montre que, dans la même expérience, ils sont capables de catégoriser les mêmes stimuli au niveau basique (fleur, chaise, humain, voiture) ou au niveau superordonné (Artificiel : voiture, chaise ou Naturel : humain, fleur) selon la consigne (Lazareva et al. 2004). Jusqu'alors cette flexibilité dans la catégorisation à différents niveaux du même objet avait été proposée comme une capacité typique, voire unique, de l'homme (Markman 1990).

Enfin, il semble que les pigeons soient capables de former des catégorisations basées sur des règles abstraites de type identité/différence dans lesquelles des paires de stimuli sont présentées, et où l'animal doit répondre si les deux objets sont identiques ou différents (Young & Wasserman 1997). Cependant, ces résultats ne sont pas aussi convaincants que dans le cas des catégories naturelles, et semblent fortement dépendre de la procédure d'apprentissage et du nombre d'essais d'entraînement.

Le pigeon possède donc un système conceptuel riche lui permettant de classer les objets du monde qui l'entoure en catégories de type basique et superordonné. Il est intéressant de noter à quel point l'hégémonie écrasante du pigeon dans ce domaine a minimisé, voire anéanti, pratiquement, toutes les recherches sur la catégorisation chez d'autres espèces d'oiseaux. Pourtant, il y a fort à parier que les corvidés qui démontrent des capacités cognitives extraordinaires dans d'autres domaines comme la mémoire épisodique, l'utilisation d'outils ou la résolution de problèmes, doivent aussi être dotés de facultés de catégorisation qui méritent d'être étudiées.

Les mammifères non primates: reconnaissance d'objets écologiquement pertinents

Les pigeons et les primates sont, indubitablement, les modèles phares pour l'étude de la catégorisation chez l'animal. Cependant, les lignées d'oiseaux et de mammifères se sont séparées il y a environ 200 millions d'années. Le cerveau d'un pigeon et celui de l'homme sont fondamentalement différents. Ainsi, il est fort probable que les facultés de catégorisation extraordinaires des pigeons, qui ressemblent tant aux nôtres, résultent d'une adaptation convergente de ces deux espèces. L'étude des capacités conceptuelles chez les mammifères non primates est donc indispensable pour retracer l'histoire phylogénétique de cette fonction si liée à la cognition générale.

On trouve assez peu de recherches sur la catégorisation chez les mammifères non primates. La raison principale en est certainement que pour nous, ainsi que pour les autres primates et les oiseaux, catégorisation rime avec vision. Or la plupart des espèces de mammifères non primates sont plus olfactives et/ou auditives que visuelles, et il est assez difficile pour nous, espèce visuelle, de concevoir un système conceptuel riche, autre que dépendant de la vision. Néanmoins, nous allons voir que vaches, chiens, dauphins, éléphants, hamsters, chevaux, phoques et moutons sont capables de former des rudiments de concepts. Rudiments, parce que les expériences menées ne remplissent pas toujours tous les critères nécessaires pour parler de « vraie » catégorisation, et qu'elles se concentrent bien souvent sur un type particulier de concept « écologique » comme celui de conspécifiques, d'apparentés, de prédateur ou de proie.

Commençons par nos compagnons de toujours, les chiens. Ils sont capables de catégoriser des photos contenant ou non un autre chien (Range et al. 2008). Après avoir été entraînés à toucher de la truffe les images contenant un chien dans des paires chiens/paysages sur écran tactile, 4 chiens sont capables de généraliser avec succès à de nouvelles images de chiens. Même s'il existe de grosses différences bas niveau entre un gros plan de tête de chien et un plan large de paysage (Torralba et Oliva 2003), les auteurs contournent habilement cette explication alternative en collant les nouveaux chiens sur les paysages familiers distracteurs.

Les vaches (prim-holstein) dans un paradigme similaire de discrimination simultanée, cette fois proposé dans un labyrinthe en Y, en s'approchant sélectivement d'une catégorie d'images associées à une récompense, sont capables de reconnaître leurs conspécifiques face à d'autres animaux domestiques (Coulon et al. 2007). Malheureusement, les auteurs n'utilisent que peu de stimuli (10 par catégorie) et ne rapportent pas de capacité de généralisation. Par contre, ils montrent que les vaches sont aussi capables de faire la tâche inverse, discriminer des non-vaches de vaches, indiquant ainsi qu'il n'y a pas de biais espèce-spécifique. En utilisant le même dispositif, la même équipe a récemment décrit que les vaches (N=8) étaient capables de reconnaissance au niveau le l'individu, qu'il soit familier ou non et de la même race ou non (Coulon et al. 2009). Cette fois, ils montrent la capacité de généralisation à des images qui n'ont pas été utilisées pendant l'entraînement et présentant divers angles de vue de la tête de l'individu. Ces résultats sont difficilement interprétables par des différences de bas-niveaux, étant donné la proximité des traits visuels entre deux individus. De plus, les vaches nécessitent moins de sessions pour catégoriser correctement les individus familiers, ce qui suggère qu'elles traitent les images comme des véritables représentations de l'individu et non comme des objets visuels arbitraires (Zayan et Vauclair 1998). Cette interprétation est renforcée par le fait que les vaches ont les oreilles baissées en arrière 10 fois plus longtemps quand elles sont face à des individus non familiers que face à des individus familiers. Ce comportement est généralement associé avec un traitement émotionnel des stimuli chez d'autres mammifères comme les chevaux et les chiens.

Mais l'espèce de bétail la plus étudiée en cognition visuelle est certainement le mouton, grâce aux nombreux travaux de l'équipe de Kendrick en Angleterre. En utilisant, comme Coulon et al. (2007) un labyrinthe en Y et un protocole de discrimination simultanée, ils sont les premiers à montrer une reconnaissance de l'individu familier de la même race, de race différente et même d'espèce différente (Kendrick et al. 1995, Tate et al. 2006). Conjointement, ils mènent des travaux de neurosciences et mettent en évidence que, comme chez l'homme, des neurones du cortex temporal (droit) codent les visages chez le mouton (Peirce et al. 2000). Néanmoins, ces démonstrations expérimentales ne présentent pas tous les

contrôles véritablement nécessaires pour montrer l'existence de "concepts" étant donné qu'il n'y pas de phase de test sur de nouveaux stimuli. Le fait que les moutons puissent retenir jusqu'à 50 associations visages d'individu/renforcements positifs ou négatifs différentes sur 2 ans (Kendrick et al. 2001) va plutôt dans le sens d'une simple discrimination et d'apprentissage d'associations par cœur. Mais la généralisation à de nouveaux stimuli avec différents angles de prise de vue a été réalisée avec succès par une étude plus récente (Ferreira et al. 2004).

Une capacité qui intéresse particulièrement les éleveurs est de savoir si leurs animaux sont capables de les reconnaître. Et effectivement, les moutons sont capables de discriminer les visages de différents humains (Peirce et al. 1998). Il en va de même pour les vaches (Taylor et Davis 1998). Ces dernières sont par ailleurs capables de reconnaître un éleveur attentionné d'un éleveur brutal (Munksgaard et al. 1997) en se basant principalement sur les vêtements qu'ils portent.

Un concept bien particulier est le concept de soi. Le fameux test du miroir de Gallup (1968) est un moyen d'évaluer si un animal/un enfant possède la conscience de soi. L'animal testé est d'abord familiarisé au miroir puis il est endormi. Pendant son sommeil, une marque est déposée sur son visage/corps. Après cette opération, si l'animal, lorsqu'il se trouve devant le miroir, présente des comportements orientés vers lui-même et en particulier vers la marque, cela signifie qu'il est conscient que l'image renvoyée par le miroir est celle de son propre reflet. Si pendant longtemps, on pensait que seuls les grands singes étaient capables de ce comportement, des expériences récentes montrent qu'il en va de même pour les éléphants (Plotnik et al. 2006), les dauphins (Reiss et Marino 2001) et les orques (Delfour et Marten 2001). Bekoff (2001) argumente pour une conscience de soi chez le chien, en pointant du doigt leur capacité à reconnaître leurs propres urines.

Comme nous l'avons souligné précédemment, la vision n'est pas le sens dominant chez bien des espèces de mammifères non primates. Existe-t-il des évidences de traitement conceptuel atteint par d'autres modalités sensorielles ? Beaucoup de mammifères possèdent des glandes spéciales qui sécrètent des substances chimiques servant de signal olfactif. Ces signaux sont des sources potentielles

pour l'identification de familiers, d'apparentés ou de partenaires sexuels et donc permettent potentiellement d'aboutir à une représentation conceptuelle de l'espèce ou de l'identité. Néanmoins, le traitement des signaux olfactifs chez les mammifères est plus souvent associé à un effet stimulus/comportement, basique et instinctif, qu'à un traitement cognitif ou conceptuel élaboré. Et pourtant, les signaux chimiques sont aussi sujets à des variations et à du bruit dont le système olfactif doit s'abstraire pour une reconnaissance efficace (Laurent 2002). Cette détection d'invariants est l'essence-même du traitement catégoriel.

Si, bien souvent, les travaux sur la reconnaissance par l'odorat s'assimilent plus à de la discrimination, quelques études ont démontré des caractéristiques indéniables de concept olfactif. Les femelles d'hamsters dorés ont plusieurs sources d'odeurs (urine, excrément, sécrétion vaginale, glande des oreilles, glandes des flancs). Dans un protocole d'habituation/ déshabituation Johnston et Peng (2008) montrent qu'un mâle hamster est capable de discriminer l'odeur de deux femelles en utilisant une de ces sources d'odeurs. En montrant qu'ils généralisent à travers les 5 types d'odeurs corporelles, les auteurs concluent que le hamster possède une représentation conceptuelle multi-composantes de l'individu. Ce processus est lié à la familiarité puisque le mâle n'est capable de former ce type de représentions multi-odeurs que s'il a été préalablement mis en contact direct avec la femelle.

L'éléphant semble aussi être capable de former des catégories olfactives. Dans le réserve d'Amboselie au Kenya, les éléphants sont confrontés à deux peuples humains différents : les Massai, éleveurs de bétail, parmi lesquels les jeunes hommes, pour montrer leur virilité ont comme coutume d'harponner les éléphants, et les Kamba, agriculteurs vivant dans des villages, qui n'ont pas spécialement d'activité nuisible aux éléphants. En utilisant des vêtements identiques portés pendant 5 jours par des jeunes hommes Kamba ou Massai, Bates et al. (2007) ont montré que les éléphants exprimaient des signes de peur accrue confrontés à l'odeur de ces derniers. Face aux vêtements portés par les Massai, ils fuient plus rapidement sur de plus longues distances, spécifiquement vers des zones comportant une végétation haute et dense (herbe à éléphant) et mettent plus de temps à revenir au calme. Il est aussi bon à savoir, pour celui qui envisagerait un

safari dans cette région d'Afrique, que les éléphants font plus souvent preuve de signes d'agressivité face à des vêtements rouges que blancs. Ces résultats sont sûrement à mettre en rapport encore une fois avec la couleur rouge des habits traditionnels Massai. Un autre comportement de discrimination étonnant constaté chez l'éléphant est celui de la reconnaissance d'ossements de congénères. Bien qu'il ne soit pas clair s'ils utilisent l'odorat, le toucher ou la vision ou une combinaison des trois, les éléphants passent plus de temps à explorer les os et particulièrement les défenses de leurs congénères morts que des os de rhinocéros ou de buffle (McComb et al. 2006).

Il existe une littérature assez abondante sur l'utilisation de cris ou vocalisations aux fonctions diverses chez les mammifères dont certaines font probablement appel à un traitement de type conceptuel. C'est par exemple le cas pour la reconnaissance du petit par la mère. Ainsi chez les phoques, la mère est capable de reconnaître le cri de son petit que ce soit dans l'eau ou sur terre, s'abstrayant ainsi des distorsions induites par le support de l'onde sonore (Perry et Renouf 1988).
La capacité d'écholocalisation des cétacés a induit une série d'expériences soulignant l'existence de représentations multi-modales d'objet (vision-écholocalisation) chez le dauphin commun. Harley et al. (1996) montrent ainsi qu'ils sont capables de réaliser une tâche d'appariement différé dans laquelle un objet perçu uniquement par écholocalisation était ensuite apparié correctement avec l'objet présenté visuellement et vice-versa.

Enfin des rudiments de concepts de type relationnel basés sur des règles abstraites ont été retrouvés chez un certain nombre de mammifères non primates. Ainsi le phoque (Mauck and Dehnhardt 2005) et le dauphin commun (Mercado et al. 2000), parmi d'autres, seraient capables de résoudre une tâche identique/différent.

Les invertébrés: des concepts dans des mini cerveaux.
J'aimerais finir ce bestiaire sur les invertébrés. Effectivement, aussi surprenant que cela puisse paraître, certains insectes et mollusques présentent clairement une capacité de catégorisation ou de reconnaissance.

Le poulpe est connu pour son « intelligence technique » qui lui permet notamment

de résoudre des problèmes relativement complexes comme celui d'ouvrir une boîte en plexiglas pour se délecter du crabe qu'elle abrite (Fiorito el al 1990). Une expérience récente montre que les poulpes passent plus de temps et ont plus de contact avec des individus inconnus qu'avec des individus auxquels ils ont été préalablement exposés (Tricarico et al. 2011). Ce phénomène - bien qu'il ne garantisse pas un traitement conceptuel de l'individu - témoigne néanmoins d'une forme de reconnaissance dont le media sensoriel reste encore à explorer.

Certains comportements découverts chez l'abeille et la guêpe remplissent parfaitement tous les critères d'une véritable catégorisation et sont eux difficiles à expliquer autrement que par un traitement conceptuel. On a du mal à imaginer que les insectes soient capables d'opérations cognitives aussi poussées. Néanmoins, les sociétés d'insectes sont d'une rare complexité, même si le fonctionnement global de la société est attribué à une intelligence collective, une série d'études récentes montre l'existence des capacités de reconnaissance et de catégorisation conceptuelle au niveau de l'individu. *Polis Fuscatus* fait partie des guêpes sociales qui vivent en sociétés organisées et hiérarchisées. Cette guêpe présente des marques jaunes sur le visage et l'abdomen d'une grande variabilité à travers la population. En modifiant artificiellement ces marques avec de la peinture, Tibetts (2002) montre que les individus ainsi maquillés subissent plus d'agressions que d'habitude par les autres membres de la colonie, suggérant l'existence d'une reconnaissance des visages chez *Polis Fuscatus*. Effectivement, des travaux plus récents de la même équipe montrent que *Polis Fuscatus* présente une capacité spécifique pour apprendre les visages de ces congénères (Sheenan et Tibetts 2011). Cet apprentissage privilégié des conspécifiques n'avait été rapporté jusqu'alors que chez les mammifères, comme l'homme (Pascalis 1998) et le mouton (Kendrick 1996). Dans le même papier, les auteurs montrent qu'une autre espèce de guêpe ne possède pas cette capacité. Ces résultats montrent à quel point des pressions sélectives de l'environnement ont pu former des adaptations similaires chez des espèces phylogénétiquement éloignées, qui ne sont pas présentes chez des espèces appartenant au même genre. Cette labilité des fonctions cognitives remet en cause sérieusement une approche purement comportementale pour l'étude phylogénétique de la cognition.

Tout aussi impressionnants sont les travaux de l'équipe de Giurfa à Toulouse sur les capacités de catégorisation visuelle chez l'abeille. Non seulement, les abeilles sont capables de former de réelles catégories basées sur des similarités physiques telle que l'orientation (Bernard et al. 2006) ou un traitement configurationnel de patterns visuels complexes (Avarguès-Weber et al. 2010), mais elles sont aussi capables de résoudre des tâches de catégorisation relationnelle basées sur des règles abstraites. Par exemple, elles sont capables de répondre différemment à des stimuli symétriques ou non et de généraliser ce concept de symétrie à la catégorisation de nouveaux stimuli (Giurfa et al. 1996). Elles sont aussi capables de résoudre la fameuse tâche similaire/différent nécessitant un concept d'identité (Giurfa et al. 2001). Un autre concept abstrait est celui de position spatiale (gauche/droite ou dessus/dessous), après avoir montré que les abeilles réussissaient la tâche dessus/dessous (Avarguès-Weber et al. 2011), Avarguès-Weber et al. (2012) viennent de publier un article où ils montrent qu'elles sont capables de combiner un concept de relation spatiale et d'identité pour résoudre une tâche de discrimination visuelle.

2.1.5 Cognition et concept chez les primates

Les premiers primates à coloniser l'Europe furent les *Dryopithèques* et les *Pliopithèques*, des singes hominoïdes principalement arboricoles mais pouvant aussi recourir occasionnellement à la bipédie, parmi eux l'oréopithèque, une sorte de grand gibbon, dont on a retrouvé des ossements en Italie. Avec l'arrivée des grandes glaciations il y a 8 millions d'années, cette lignée va disparaître d'Europe, non sans que certaines espèces se soient préalablement déplacées en Asie. Ce n'est que 6 millions d'années plus tard que le pied d'un autre primate foule le sol européen. Il s'agit cette fois d'un représentant du genre Homo, l'Homo erectus dont les plus vieux ossements on été retrouvés en Espagne et datent d'1,8 million d'années. Depuis, hormis une petite colonie de macaques magots installés sur le rocher de Gibraltar, nous sommes la seule espèce de primates vivant sur le continent européen. En l'absence de nos plus proches cousins, ce n'est donc pas étonnant que la tradition intellectuelle, philosophique et religieuse occidentale ait mis une telle distance entre l'animal et l'homme. Il y a fort à parier que si

Descartes avait pu observer un groupe de bonobos interagir, il aurait modéré ses thèses sur l'homme doué de raison et l'animal-machine. Sans surprise, le premier scientifique à s'intéresser vraiment aux singes et à leur cognition est Charles Darwin. En 1838, bien avant la publication de *The origin of Species,* il écrivait dans un cahier de notes « Celui qui comprendra les babouins apportera plus à la métaphysique que Locke » (Notebook M84e). Il faut croire qu'encore une fois les propos de Darwin n'ont pas été des paroles en l'air puisque depuis lors, de nombreux travaux ont été menés pour comprendre les autres primates, les primates non humains.

Historique

Les premiers comportementalistes à se lancer dans le défi proposé par Darwin et à se pencher vraiment sur la psychologie des singes sont Wolfang Köhler et Robert Yerkes. Dans son laboratoire de Tenerife sur les îles Canaries, Köhler, va poser des problèmes variés à son groupe de 9 chimpanzés. Lorsqu'il attache une banane trop haut pour que les animaux puissent l'attraper et qu'il laisse des caisses en bois à leur disposition dans l'enclos, les singes vont directement empiler les caisses nécessaires pour se saisir du fruit convoité. De même, ils sont capables de réaliser des emboîtements complexes de tuyaux pour former une perche, suffisamment grande pour leur permettre de faire rouler jusqu'à eux, une pomme déposée à plusieurs mètres de leur parc. Les singes n'utilisent donc pas une stratégie d'essai-erreur nécessitant de tester absolument toutes les possibilités afin d'atteindre leur but, au contraire comme l'homme, ils ont une compréhension soudaine du problème que Koehler appelle « insight ». C'est cet « insight » qui leur permet en un essai de réaliser le dispositif adéquat pour obtenir la récompense. Bien que ces expériences aient été réalisées entre 1913 et 1917, en pleine expansion behaviouriste, de tels comportements laissent peu de place à une explication autre que l'existence de représentations chez le chimpanzé, lui permettant de manipuler mentalement les objets pour trouver la bonne combinaison et ensuite seulement de la mettre en œuvre matériellement.

A la même période, aux Etats-Unis à Atlanta, Yerkes fondait un laboratoire (qui existe toujours et qui porte maintenant son nom), dédié à l'étude du comportement des primates non humains. Le psychologue américain aborde des sujets aussi

variés que la mémoire de travail spatiale ou la coopération. C'est à son initiative que les premiers grands singes sont élevés comme des enfants pour étudier l'impact de cette éducation sur leurs capacités cognitives et de communication (Kellogg et Kellogg 1933, Hayes et Hayes 1951).

Néanmoins ces travaux pionniers en primatologie, déjà cognitifs, vont rester dans l'ombre du behaviourisme despotique de l'époque pas encore prêt à concéder la moindre once de représentation à l'animal. C'est donc avec une approche encore très behaviouriste, et à travers le conditionnement opérant de Skinner, que 30 ans plus tard, à la fin des années 60, Harry Harlow mais aussi Duane Rumbaugh s'intéressent aux capacités d'apprentissage des primates. En apprenant aux singes à associer une récompense à un type d'objet, les recherches de l'époque commencent à se pencher sur les capacités de discrimination des animaux. Le souci majeur de ces travaux est d'évaluer la flexibilité de l'apprentissage, ce qu'on appelle aussi l'apprentissage à apprendre. Les expérimentateurs se rendent ainsi compte que si les primates non humains ont été préalablement entraînés à associer, au sein d'une paire d'objet, l'un d'entre eux à une récompense, ils sont capables de transférer cette règle à une autre paire d'objets, en moins d'essais qu'il n'en a fallu pour apprendre la tâche originale. On doit à ces études sur "l'apprentissage à apprendre" des travaux d'une rare exhaustivité en matière de cognition comparée qui, comme l'illustre la Figure 2, permettent la comparaison de plus d'une douzaines d'espèces de primates et semblent indiquer une hiérarchie depuis les prosimiens jusqu'aux grands singes en matière de flexibilité d'apprentissage.

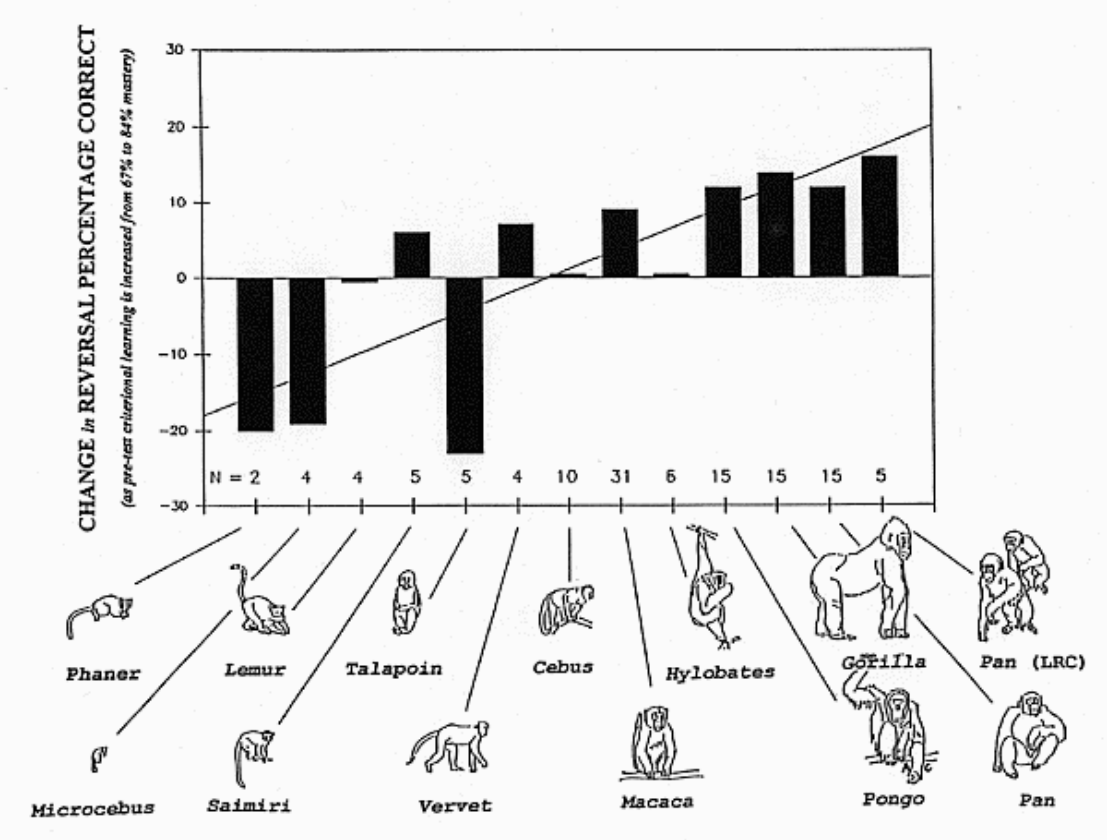

Figure 2: Capacité de transfert d'apprentissage pour 13 espèces de primates non humains (from Rumbaugh et Pate 1984)

Les premières études sur les comportements naturels des grands singes commencent aussi vers la fin des années soixante. Jane Goodall après des années d'observation publie en 1970 la première utilisation d'outils chez les chimpanzés (Goodall 1970).

En parallèle, la révolution cognitive fait rage chez les psychologues et les idées behaviouristes sont rapidement éliminées de toute approche visant à étudier les comportements humains. Le transfert de la notion de cognition aux animaux non humains mettra lui un peu plus de temps à se mettre en place. Néanmoins, l'idée selon laquelle les animaux peuvent eux aussi former des représentions mentales va décomplexer les esprits et permettre de découvrir l'existence de capacités cognitives inespérées chez nos cousins primates, notamment les grands singes. En promenant de jeunes chimpanzés sur ses épaules alors qu'il cache de la nourriture dans 18 endroits de leur enclos, Emile Menzel (1973) montre que ces animaux possèdent une excellente mémoire de travail spatiale. Effectivement, quand dans une phase de test, les animaux sont relâchés dans l'enclos, ils se rappellent la

position et le type de nourriture de chaque cache. De plus, pour visiter les emplacements ils utilisent un trajet pratiquement optimal, indépendant du trajet du cacheur, et revisite rarement deux fois une même cache. De tels comportements suggèrent fortement l'utilisation d'une carte cognitive spatiale. David Premack se lance dans l'étude des capacités mathématiques et d'une « théorie de l'esprit » (capacité à appréhender les intentions d'autrui) chez les chimpanzés élevés par l'homme (Premack et Woodruff 1978). Deux études montrent que les grands singes sont capables d'apprendre un langage complexe, en utilisant le langage des signes (Gardner et Gardner 1969) ou en se servant de lexigrammes (symboles visuels représentant des mots) (Savage-Rumbaugh et al. 1978 a). On découvre aussi que les macaques possèdent des comportements alimentaires culturels (Kawai 1965) et qu'il existe dans les sociétés de chimpanzés une véritable « politique » (de Wall 2007).

Depuis ces études pionnières, l'approche cognitive en primatologie a connu un véritable engouement et des centaines d'études ont été menées chez de nombreuses espèces. Nous allons nous concentrer dans la partie suivante sur les capacités catégorielles et conceptuelles qui font l'objet de cette thèse.

Concepts naturels chez les primates non humains

De manière assez surprenante, la littérature sur les capacités de catégorisation chez les primates non-humains est moins importante que celle qu'on peut trouver chez le pigeon. En revanche, les types de protocoles et d'approches sont eux beaucoup plus variés. Un des plus beaux exemples d'existence de concepts naturels chez les primates nous vient de l'éthologie cognitive et de l'étude de terrain mené par Seyfarth, Cheney et Marler sur les singes vervet, au Kenya dans la réserve d'Ambosélie (Seyfarth et al. 1980). Les éthologues se sont d'abord rendus compte que les singes possédaient plusieurs cris d'alarme associés à différents types de prédateurs. En enregistrant ces cris d'alarmes et en les repassant en « playback », ils ont montré que les singes présentaient une réaction différente en fonction du type de cris. Si le cri correspond à un léopard alors ils se dépêchent de rejoindre les arbres les plus proches pour s'y percher, si le cri correspond à la présence d'un aigle tous les regards s'orientent vers le ciel, au contraire si le cri correspond à la présence d'un serpent, les singes se mettent à scruter le sol. Ces comportements

fascinants montrent l'existence de communications sémantiques documentant la catégorie de prédateur pour l'émetteur et les receveurs. Cette capacité semble fortement liée à l'expérience. Les très jeunes singes produisent des cris pour la plupart des oiseaux, les juvéniles n'alarment déjà plus que pour les rapaces et les adultes eux ne signalent que l'aigle martial, leur principal prédateur aérien.

Evidement, à l'instar de ce qui a été fait chez le pigeon, les primates non humains ont aussi été engagés dans des tâches de catégorisation par conditionnement opérant. D'Amato et Van Sant (1988) démontrent ainsi que les capucins moines sont capables de catégoriser des photos comme contenant un humain ou non. Néanmoins, en s'intéressant en détail aux erreurs produites par les animaux sur les images ne contenant pas d'humains, les auteurs se rendent compte qu'elles contiennent souvent une zone rougeâtre, rappelant celle d'un visage humain. Il semblerait donc que les capucins n'aient pas réellement formé un concept d'humain mais aient choisi la stratégie alternative de répondre sur un indice visuel bas niveau co-occurrent leur permettant de résoudre la tâche et d'obtenir leur récompense. Cette étude illustre le danger d'utiliser des photographies pour l'étude des concepts naturels et souligne le soin qu'il faut porter aux stimuli et aux analyses pour exclure l'utilisation d'indices visuels bas niveau.

La même année, dans une étude comparative, Roberts et Mazmanian (1988) s'intéressent aux capacités des pigeons, des humains et des singes écureuils à former des catégories naturelles répondant aux trois niveaux d'abstraction définis par Rosch. Au niveau le plus concret, ou subordonné, les sujets doivent répondre sur un oiseau particulier : le martin pêcheur face à des images d'oiseaux d'autres espèces. Au niveau intermédiaire, ou basique, ils doivent classer les oiseaux parmi des animaux non-oiseaux. Finalement au niveau le plus abstrait, ou superordonné, les sujets doivent former une catégorie animale parmi des stimuli non animaux. Après un entraînement d'une durée variable pour chaque espèce, les performances des sujets sont évaluées dans une phase de transfert à de nouveaux stimuli. Les hommes réalisent la tâche sans difficulté aux trois niveaux d'abstraction et, conformément aux résultats de Rosch, atteignent leurs meilleures performances pour le niveau basique. La tâche qu'ils réussissent le moins bien est celle de la catégorisation subordonnée. Les pigeons et les singes montrent le profil inverse:

leurs meilleures performances sont enregistrées au niveau concret et ils ne dépassent pas le niveau chance lorsqu'il s'agit de généraliser au niveau basique. De manière surprenante, malgré le fort niveau d'abstraction que présente le niveau superordonné, de par la variabilité des exemplaires qui le composent, les singes sont aussi capables de former assez rapidement un concept d'animal. Les pigeons mettent plus de temps que les singes mais réussissent aussi correctement cette catégorisation superordonnée. Cependant les différences de performances entre les hommes et les deux espèces d'animaux sont telles qu'elles laissent présager l'existence de mécanismes neuronaux et cognitifs divergents même entre les 2 espèces de primates.

Michèle Fabre-Thorpe et son équipe de Toulouse, ont voulu creuser la capacité des singes à former des catégories naturelles au plus fort degré d'abstraction qu'est le niveau superordonné, à travers une série d'études publiées à la fin des années 90 et au début des années 2000 (Fabre-Thorpe 2003). Elle note que, dans l'étude de Roberts et Mazmanian (1988), les performances pour les singes écureuils au niveau superordonné, bien que significativement meilleures, sont proches du niveau chance, dépendent de la taille du set d'entraînement et surtout diffèrent considérablement des performances des humains. Elle attribue ces faibles performances à plusieurs facteurs. D'abord, en étudiant les singes écureuils qui sont des singes du nouveau monde et qui ont divergé des singes de l'ancien monde il y a environ 30-40 millions d'années, on a plus de chance de trouver des capacités cognitives différentes de celles des humains qu'en utilisant des primates non humains qui nous sont plus proches. Fabre-Thorpe et al. vont donc s'intéresser aux macaques rhésus, une espèce de singe du nouveau monde dont la lignée a divergé de celle de l'homme il y a environ 25 millions d'années. Roberts et Mazmanian utilisent un protocole de choix forcé où deux stimuli sont présentés simultanément ce qui encourage les sujets à utiliser une stratégie de discrimination entre ces deux exemplaires particuliers plutôt que de recourir à l'utilisation d'une représentation interne de la catégorie comme dans la procédure de go-no go où les images sont présentées une à une. Fabre-Thorpe critique aussi la distinction nette entre phases d'entraînement et phases de test. En entraînant de manière intensive les animaux sur un faible nombre d'exemplaires, ils sont poussés à mémoriser ces exemplaires particuliers plutôt qu'à former un concept général d'animal. De plus,

dans la phase de test, pendant laquelle les sujets sont exclusivement confrontés à des stimuli nouveaux, ils peuvent plus facilement fluctuer dans leur état cognitif car la moindre erreur initiale va troubler l'animal en le privant de récompense et, en impactant sa concentration et son attention, va accentuer petit à petit le nombre d'erreurs. Dans la comparaison entre l'homme et le singe, Fabre-Thorpe note le faible nombre d'exemplaires d'images d'animaux proposées aux singes durant l'entraînement comparé aux milliers de confrontations que chaque homme a pu avoir au cours de sa vie avec cette catégorie sous une forme ou une autre. Finalement, Fabre-Thorpe pointe aussi du doigt le temps de présentation des images et le temps laissé au sujet pour répondre. Effectivement, si singes et hommes possèdent initialement un mécanisme cognitif commun permettant de réaliser une catégorisation superordonnée, il est possible que chez l'homme d'autres phénomènes cognitifs liés au langage ou à l'éducation puissent se surajouter à ces mécanismes de base lorsqu'ils ont suffisamment de temps pour réaliser la tâche.

Afin de pallier tous ces handicaps Fabre-Thorpe et al. (1998) vont adapter aux singes une tâche qui venait alors d'être développée chez l'homme (Thorpe et al. 1996) : la tâche de catégorisation rapide en go-no go. Dans l'étude originale, les sujets humains sont assis face à un écran d'ordinateur, des images leur sont flashées pendant 20 ms. Leur tâche est de presser une touche pour faire défiler les images et de lever le doigt aussi vite que possible s'ils aperçoivent un animal dans la scène. En analysant les temps de réaction et le signal EEG conjointement enregistré, les auteurs documentent la formidable efficacité du système visuel humain qui permet de résoudre une tâche de catégorisation abstraite à partir de stimuli complexes en un temps extrêmement court.

Fabre-Thorpe et al. (1998) engagent 2 singes macaques rhésus dans un protocole extrêmement similaire avec les mêmes stimuli. La seule différence est la récompense que les singes obtiennent après chaque bonne réponse. Pour éviter l'apprentissage par cœur du set d'entraînement et pallier le manque d'expérience des animaux avec les photographies, des stimuli nouveaux sont introduits tous les jours dans une faible proportion par rapport aux stimuli familiers et un très large jeu d'images est utilisé. Enfin, pour montrer que la capacité à former des concepts

superordonnés ne s'arrête pas uniquement à la catégorie "animal", un des singes est entraîné à faire une catégorisation aliments/ non aliments. Quelle que soit la catégorie, les singes sont extrêmement bons pour réaliser la tâche. Ils atteignent 90% de réussite pour catégoriser des images qu'ils n'ont jamais vues auparavant (Fabre-Thorpe et al. 1998, Delorme et al. 2000). Ces performances sont directement comparables avec celles des humains qui avoisinent les 94% correct (Thorpe et al. 1996). De manière assez surprenante les singes sont même beaucoup plus rapides que les humains pour réaliser cette tâche complexe. Alors que les humains ont un temps de réaction médian de 370 à 450 ms, celui des singes ne dépasse pas les 300 ms et descend même jusqu'à 250 ms. Pour évaluer le temps de réaction minimal permettant de réaliser cette catégorisation, Thorpe et al. définissent une mesure spéciale appelée temps de réaction minimal. C'est le temps minimum à partir duquel les sujets produisent significativement plus de bonnes que de mauvaises réponses. Même avec cette mesure extrêmement contraignante d'un point de vue temporel, les singes avec un temps de réaction minimal de 180-200 ms gardent une avance incompressible de 50-70 ms par rapport aux humains dont le temps de réaction minimal est de 250-270 ms.

La première explication qui vient à l'esprit pour analyser ces différences est celle de l'utilisation de caractéristiques visuelles bas niveau par les singes, comme dans l'étude de D'Amato et Van Sant (1988), plus rapidement traitées que celle nécessaires pour atteindre le concept d'animal utilisé par les hommes. Afin de tester cette hypothèse Delorme et al. (2000) testent la capacité des singes à réaliser la même tâche sur des stimuli en noir et blanc : les performances sont inchangées pour les hommes comme pour les singes. Si les singes utilisent un indice bas niveau ce n'est donc pas la couleur. Dans une autre étude comparative, Macé et al, en faisant varier le niveau de contraste dans les images caractérisent un peu plus les représentations utilisées par les hommes et les singes (Macé et al. 2010). Les deux primates sont encore capables de réaliser la tâche même avec un contraste très faible et leurs performances sont affectées de manière similaire par les différents niveaux de contraste. Les auteurs concluent que les représentations mises en jeu pour cette catégorisation rapide seraient assez rudimentaires et pourraient être supportées par la voie visuelle rapide, dite magnocellulaire, achromatique et robuste aux faibles contrastes.

De plus, dans une étude chez l'homme où la tâche est rendue particulièrement difficile, en utilisant un pattern contrasté masquant l'image 10 à 20 ms après sa présentation (Bacon-Macé et al. 2005), les sujets réussissent la tâche mais disent ne pas être pas confiants dans la réponse donnée, comme si la représentation mise en jeu suffisait pour déclencher une réponse catégorielle rapide mais n'induisait pas forcément une perception consciente.

En poussant le sujet à répondre le plus vite possible, la tâche de catégorisation rapide se libère des processus cognitifs parasites et permet d'accéder à l'essence des mécanismes cognitifs et neuronaux suffisant à l'élaboration d'un concept visuel même abstrait. De plus, les conditions expérimentales extrêmement similaires utilisées chez l'homme et le singe en font un outil de choix pour la cognition comparée. C'est donc celui auquel nous aurons recours dans notre recherche d'un système homologue de conceptualisation visuelle chez le primate.

2.2 Article1 : Humains et macaques rhésus partagent des représentations visuelles haut niveau

Problématique

La capacité à former des concepts, notamment naturels ou écologiques, semble être une capacité extrêmement répandue au sein du règne animal. Etant donné le nombre d'études mettant en valeur ces fonctions conceptuelles chez des espèces aussi variées que l'abeille, le pigeon, le poulpe, la vache ou le singe écureuil, il y a tout à parier que la plupart des animaux sont capables d'accéder à certains niveaux d'abstraction. Former des catégories leur permet de reconnaître leurs prédateurs, leurs proies, leurs partenaires sexuels ou encore les individus qui leur sont proches. Selon mon opinion, les espèces parmi lesquelles aucune capacité de reconnaissance ou de catégorisation n'ont été diagnostiquées, soit n'ont pas encore été étudiées, soit elles l'ont été avec de mauvais outils, des questions non pertinentes ou des protocoles non adéquats. Il n'est donc plus de notre ressort d'essayer de convaincre de la capacité des animaux à former des représentations mentales. Les animaux sont capables d'élaborer des concepts riches et par conséquent accèdent à une certaine forme de pensée. Néanmoins, nous avons aussi vu qu'une espèce de guêpe, dont l'anatomie et la physiologie cérébrale sont si différentes des nôtres, présente des capacités d'apprentissage spécifiques des visages de leur conspécifiques, tout à fait comparables à celles de l'être humain. Un comportement qui n'a pas été retrouvé chez une autre espèce de guêpe du même genre. De tels résultats mettent évidemment en évidence l'existence de fonctions d'abstraction convergentes et non dérivées d'un organe commun, mais surtout ils posent la question de la labilité des fonctions cognitives. Si des espèces très éloignées (comme l'homme et la guêpe) d'un point de vue phylogénétique, par des critères morphologiques et génétiques, possèdent des fonctions cognitives très similaires, qui ne sont pas retrouvées dans une espèce beaucoup plus proche (une autre espèce de guêpe) alors on est en droit de douter de l'existence d'une réelle phylogénie de la cognition. Est-ce que la notion de continuité et d'évolution biologique des espèces, prônée par Charles Darwin, ne s'appliquerait en fait pas aux fonctions cognitives? Est-ce que chaque espèce formerait son propre répertoire cognitif indépendamment des espèces qui lui sont les plus proches phylogénétiquement? En particulier est-ce que nous « êtres humains » partageons

des mécanismes cognitifs avec d'autres espèces animales? Où est-ce que cette pensée que nous expérimentons à chaque seconde nous est absolument propre? Les réponses à ces questions se trouvent dans la comparaison la plus méticuleuse possible entre les capacités cognitives des espèces.

Nous avons parlé précédemment de la tâche de catégorisation rapide qui permet de saisir le rudiment de représentation suffisant pour former un concept visuel à la fois chez l'homme et le macaque Rhésus, dans des conditions expérimentales extrêmement identiques. C'est comme nous l'avons expliqué précédemment, un outil idéal pour comparer la capacité des deux primates à former le concept d'une catégorie naturelle. Néanmoins, nous avons aussi expliqué que l'utilisation de scènes naturelles pouvait amener les animaux à utiliser des stratégies alternatives aux nôtres en se fiant à des indices visuels bas niveau. Delorme et al. (2000) ont exclu l'utilisation des couleurs, mais depuis les années 2000 la vision par ordinateur a considérablement évolué et on s'est aperçu que d'autres caractéristiques des images (Torralba et Oliva 2003), plus subtiles, pouvaient aussi permettre de déterminer la catégorie globale d'une scène (naturelle, artificielle, mer, montagne). De plus, ces statistiques de l'image permettent de prédire la présence d'une catégorie d'objets particulière (animaux, voitures…) avec des performances approchant les 80% de précision. La Figure 3 illustre à quel point l'image moyenne (moyenne de la luminance des pixels) et la signature spectral (moyenne des spectres d'amplitudes) sont différents pour des scènes naturelles contenant un animal ou une voiture.

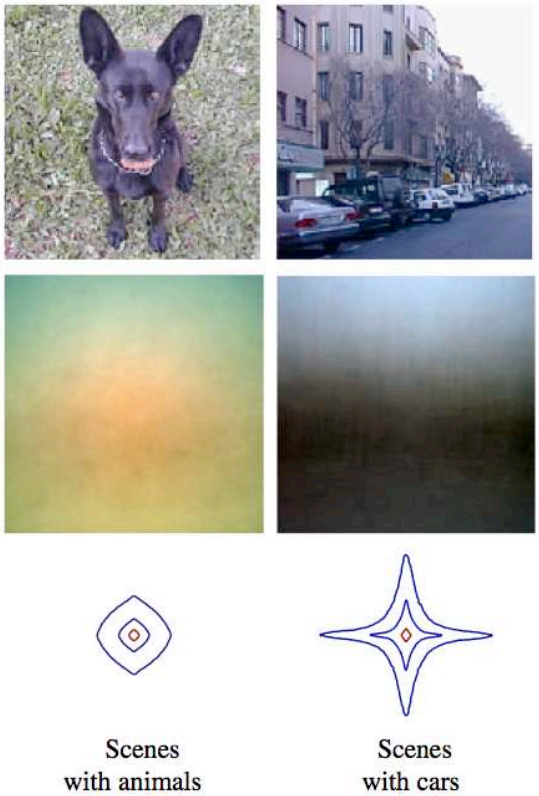

Figure 3: Image moyenne (au milieu) et signature spectrale (en bas, les contours représentes depuis l'intérieur vers l'extérieur 60% 80% et 90% de l'énergie du spectre d'amplitude) des images naturelles contenant un animal (à gauche) ou une voiture (à droite)

Or, depuis les premières expériences menées en 1998, l'équipe de Fabre-Thorpe a toujours utilisé les mêmes singes qui ont appris pendant des années à catégoriser des photographies issues de la même banque d'images (Corel et photographies personnelles). Dans cette banque d'images, la grande majorité des images d'animaux utilisés comme cibles pour les singes, présentent l'animal photographié dans son milieu naturel, alors que bien souvent les images utilisées comme distracteur présentent des scènes ou des objets artificiels. De telles différences, sont extrêmement bien captées par le modèle bas-niveau de Torralba et Oliva (2003). Avec le temps, les singes pourraient très bien avoir appris la co-occurrence de ces informations contextuelles bas-niveau, et les utiliser pour résoudre la tâche de catégorisation qui leur est proposée, à l'instar du modèle computationnel. Ces informations bas-niveau extraites dès le cortex visuel primaire, pourraient expliquer la rapidité des singes par rapport à celle des humains.

Dans une tâche de catégorisation à partir de photographies naturelles, est-ce que le singe utilise des statistiques globales bas niveau ou est-ce qu'il se focalise vraiment

sur les traits visuel de l'objet à catégoriser ? A quel point les représentations mises en jeux par l'homme et le singe dans cette tâche de catégorisation rapide sont-elles similaires ? Voilà les questions auxquelles cet article cherche à répondre.

2.2.1 Résumé de l'article

L'astuce expérimentale mise en place dans cet article est l'élaboration de quadruplets de stimuli (2 objets: un animal et un objet manufacturé et 2 contextes: un naturel et un artificiel, **Figure 4**). En collant les objets, préalablement égalisés en surface et en indices visuels bas niveau, exactement à la même position dans les deux contextes, on oblige le sujet à se concentrer sur l'objet et à faire abstraction de l'information contextuelle.

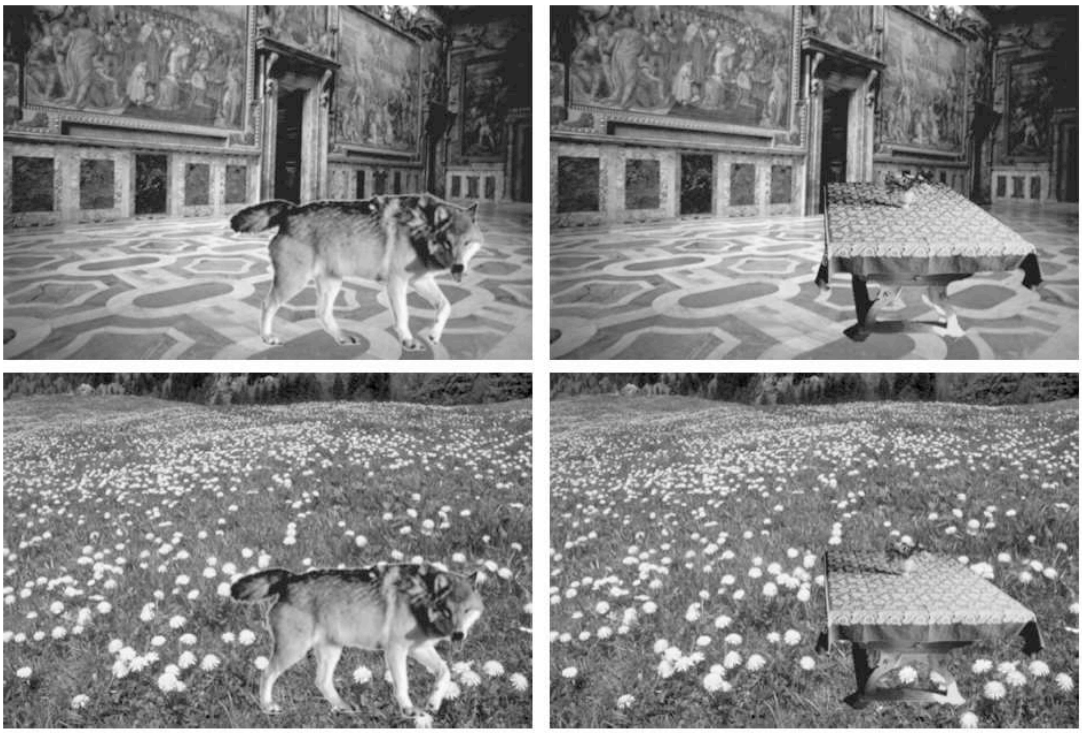

Figure 4 : Un exemple de quadruplet. Chacun des 192 appariements contient 4 stimuli: 2 stimuli congruents (bas gauche et haut droite), 2 stimuli incongruents (haut gauche et bas droite) qui sont aussi: 2 stimuli cibles (première colonne), 2 stimuli distracteurs (deuxième colonne). Les quatre images ont la même luminance moyenne et le même contraste RMS. Les quatre objets (2 loups et 2 tables) ont la même surface, la même position dans l'image (excentricité), la même luminance moyenne et le même contraste RMS. Ces deux derniers paramètres tiennent compte de la luminance et du contraste des fonds sur lesquels ils sont collés, pour préserver les contrastes locaux objet/contexte.

Comme nous l'avons dit précédemment, les singes ont été entraînés pendant des années à catégoriser des images naturelles associant bien souvent animaux cibles et fond naturel et non-animaux distracteurs et fond artificiel. Néanmoins, dès la première présentation de nos stimuli manipulés, à l'instar des hommes, les réponses go vers les cibles sont plus nombreuses sur les stimuli contenant un animal que sur des stimuli au contexte strictement identique, mais dans lesquelles l'animal a été remplacé par un objet. Ces résultats sont vrais avec un fond naturel comme avec un fond artificiel. Comme les hommes, nos deux singes, pour résoudre la tâche de catégorisation animal/non animal utilisent l'objet et sont capables de le faire quelle que soit la nature de l'information contextuelle.

Cependant, il semble que même chez l'homme ces performances de catégorisation rapide soient légèrement biaisées par la relation contexte/objet au niveau superordonné (Joubert et al. 2008). On retrouve cet effet dans notre groupe humain contrôle qui, sur les trois séances qu'ils réalisent, sont toujours plus précis et plus rapides pour catégoriser des stimuli présentant une association objet/contexte congruente (animal sur fond naturel ou objet manufacturé sur fond artificiel) que des stimuli incongruents (animal sur fond artificiel et objet manufacturé sur fond naturel). Les singes présentent des effets extrêmement comparables en précision, temps de réaction minimal et robustesse de l'effet à travers les sessions. Pour les deux espèces de primates cet effet est d'autant plus fort que les objets sont de petite taille. Par contre sur les objets de grosse taille les performances globales augmentent et l'effet objet/contexte disparaît. En contraste, un modèle de vision par ordinateur réputé pour donner de bons résultats dans la reconnaissance de scènes naturelles et même d'objets (Torralba et Oliva 2003) est tout à fait incapable de réaliser la catégorisation animal/non animal sur notre set de stimuli, même lorsque les objets sont gros.

ARTICLE 1: Humans and Monkeys share visual representation

Denis Fize, Maxime Cauchoix, Michèle Fabre-Thorpe

(2011), PNAS, 108(18), 7635-40

Humans and monkeys share visual representations

Denis Fize[a,b,1], Maxime Cauchoix[a,b], and Michèle Fabre-Thorpe[a,b]

[a]Université de Toulouse, Université Paul Sabatier, Centre de Recherche Cerveau et Cognition, 31062 Toulouse, France; and [b]Centre National de la Recherche Scientifique, Centre de Recherche Cerveau et Cognition, Faculté de Médecine de Rangueil, 31062 Toulouse, France

Edited by Robert Treichler, Kent State University, Kent, OH, and accepted by the Editorial Board March 23, 2011 (received for review November 3, 2010)

Conceptual abilities in animals have been shown at several levels of abstraction, but it is unclear whether the analogy with humans results from convergent evolution or from shared brain mechanisms inherited from a common origin. Macaque monkeys can access "non-similarity–based concepts," such as when sorting pictures containing a superordinate target category (animal, tree, etc.) among other scenes. However, such performances could result from low-level visual processing based on learned regularities of the photographs, such as for scene categorization by artificial systems. By using pictures of man-made objects or animals embedded in man-made or natural contexts, the present study clearly establishes that macaque monkeys based their categorical decision on the presence of the animal targets regardless of the scene backgrounds. However, as is found with humans, monkeys performed better with categorically congruent object/context associations, especially when small object sizes favored background information. The accuracy improvements and the response-speed gains attributable to superordinate category congruency in monkeys were strikingly similar to those of human subjects tested with the same task and stimuli. These results suggest analogous processing of visual information during the activation of abstract representations in both humans and monkeys; they imply a large overlap between superordinate visual representations in humans and macaques as well as the implicit use of experienced associations between object and context.

homology | natural scenes | non-human primate | visual categorization

In a demonstration of monkeys' abstraction abilities, Bovet and Vauclair (1) reported that monkeys were able to classify new pairs of real objects as "same" when belonging to the same superordinate category (an apple and a banana, or a cup and a padlock), while classifying other combinations as "different." The ability to perform a judgment of conceptual identity among categories such as Food or Tools corresponds to an abstract level of conceptualization (2–4), but the nature of the cerebral processes, the mental representations involved, and their similarity with those of humans remain unclear.

In monkeys, the ability to form and access perceptual classes is the most extensively studied abstraction level. Categorizing by perceptual similarity enables the formation of open-ended categories, which generalize to novel elements of the same kind (2, 3). This categorical behavior emerges spontaneously even in macaque monkeys with lesions of the lateral prefrontal cortex (5), a structure in which neurons show category-specific selectivity (6–8). In fact, building "perceptual concepts" (4) such as natural visual categories—trees, cats, dogs, etc. (6, 9)—or artificial ones such as letter categories (10, 11) could rely on relatively simple mechanisms: a set of representative visual features involving particular shapes or typical shadings could be used as diagnostic information to categorize objects at this perceptual level (12). Such processing in humans and monkeys is likely to take place in the inferior temporal cortex, which has long been known to be critical for object visual recognition (13, 14) and generalization of object views (15). However, single-neuron activity within this region does not seem to reflect categorical information (9, 16–18).

At a more abstract level, conceptual behavior implies categorizing beyond the physical similarity between exemplars of a class (3, 4). The few behavioral studies that have investigated this abstraction level in monkeys dramatically increased stimulus variety by using pictures of natural scenes and superordinate categories such as Food or Animal to avoid diagnosticity from a restrained set of low-level visual cues (1, 19–22). In the Animal category, for example, the large variety of perceptually different instances of mammals, birds, insects, or fishes and the high performance reached on new exemplars suggest an abstract level of categorization and an ability for macaque monkeys to access superordinate representations (23).

However, the claim that abstract representations are used to perform the superordinate categorizations was challenged by the finding that global scene statistics can predict the presence of animals or objects in scene pictures (24). Indeed, to succeed in such tasks, monkeys had been extensively trained by using large image sets extracted from commercial databases (19–21). Because the animal images are much more likely to be pictured on natural scene backgrounds than in urban contexts, the scene statistics that support the distinction between natural and man-made environments (25, 26) and the presence of foreground objects could potentially allow performance above chance. Monkeys could have used such contextual regularities to reach high scores without any conceptual representation of the object category.

Here, we adapted for monkeys the approach used in earlier work investigating contextual effects on object processing during visual scene categorization by humans (27): animal target objects and man-made distractor objects were displayed randomly in either man-made or natural scene backgrounds. Using such stimuli should confuse the monkeys if they rely on global scene statistics to perform the task. Possible similarities between the mental representations used by humans and monkeys to solve such visual categorization tasks were also investigated by using the categorical object/context interference phenomenon. Indeed, in humans, categorization performance is affected in terms of both delayed reaction times and lower accuracies when the object and the background context belong to different superordinate categories (27–31), suggesting that the mental representations of object and background context overlapped during task performance and were "category-sensitive" at the superordinate level (27, 31).

Results show that monkeys' categorization performance primarily relied on the processing of animal/object information, and that the scene background did not play a major diagnostic role. Furthermore, the categorical interference between foreground objects and background contexts had very similar effects on monkey and human performance over a wide range of parameters. These data suggest a high analogy between monkeys and humans for the use of visual cues to access categorical representation and their cerebral processing.

Results and Discussion

Two macaque monkeys (Dy and Rx) and 11 human subjects performed the Animal/Non-Animal rapid scene-categorization task developed by Thorpe et al. (32) and adapted for monkeys by Fabre-Thorpe et al. (20). The response consisted of a button re-

Author contributions: D.F., M.C., and M.F.-T. designed research; D.F. and M.C. performed research; M.C. analyzed data; and D.F. and M.F.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.T. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: denis.fize@cerco.ups-tlse.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016213108/-/DCSupplemental.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

lease and a screen touch performed in under 1 s when a briefly flashed stimulus (50 ms) contained an animal (go/no-go task; Fig. 1*A*). The test stimuli were composed of 192 associations of four achromatic scenes counterbalancing superordinate context categories (man-made and natural) and superordinate object categories (animal or man-made object): very varied animal and man-made object exemplars were presented equally often in the natural and man-made background contexts. Test stimuli were only seen once by each subject, and backgrounds and objects were controlled for various low-level visual characteristics (*Materials and Methods* and Fig. 1*B*).

For monkeys, the sequence of stimuli randomly interleaved test stimuli with familiar scenes to ensure the stability of monkey motivation and performance (33); first-trial responses to the test stimuli were thus of crucial importance. Familiar scenes were taken from the commercial photographs on which monkeys initially learned the Animal/Non-Animal task and had been intermittently trained and tested (for 5–6 y) (*Materials and Methods*); both monkeys performed at 89% correct on these familiar stimuli during the days preceding the present experiment.

Non-Similarity–Based Concepts in Macaque. For both monkeys, first-trial performance on the test stimuli showed accuracy scores significantly above chance, regardless of the nature of the stimuli background (Fig. 1). The monkeys were able to use both man-made objects and animals as the pertinent key target for task performance: the proportion of hits outnumbered very significantly the proportion of false alarms in both monkeys (Fig. 1*C*). Monkeys were also able to ignore the scene background category because global-task accuracy was well above chance level, regardless of the environment category (Fig. 1*D*): natural (Rx, 71%; Dy, 67%) or man-made (Rx, 69%; Dy, 63%).

This task was particularly hard to perform given the high ratio of test stimuli introduced among familiar pictures (1/3 test, 2/3 familiar stimuli). In addition, in the test stimulus set, objects averaged only 6% of scene surface (range 0.2–22%), a much smaller object/scene surface range than in familiar stimuli. Object locations in the test stimuli were also extremely varied at eccentricities ranging from 0.2° to 24° from fixation (12° on average), which probably accounts for the conservative strategy exhibited by both monkeys (there were more misses than false alarms: both monkeys, χ^2 tests, $P < 10 \times 10^{-5}$) and the accuracy drop compared with previous studies of our group [usually 90% correct (20, 21)]. In fact, a similar drop in performance accuracy was also observed in the 11 human subjects performing the same task with the same test stimuli: they averaged only 80% correct, a performance that has to be compared with the 94% correct usually reached by humans on novel scenes (19, 32, 34).

Thus, despite their long training with commercial photographs that mostly associate animals with natural backgrounds, the two monkeys were clearly able to use object information and to ignore background information when categorizing new scenes. From the very first presentation of the manipulated stimuli, the bias attributable to the scene background category only accounted for ~2–5% of the global accuracy (Fig. 1*D*), despite the relatively small size of most objects within the scenes. Such immediate generalization for new man-made objects and animal exemplars presented in unusual scene contexts rules out the possibility that scene background regularities alone could explain performance.

Such results obtained in monkeys show that superordinate representations supporting abstract concepts do not necessarily require high-level functions such as linguistic abilities or even elaborate processing far from perceptual modalities (23, 35, 36). Recent results indicate that superordinate categories are the first to be accessed within the visual modality (37, 38). In fact, several processing schemes have been proposed that combine such coarse-to-fine visual processing (26, 39–41) with fast decision mechanisms. For example, coarse visual information could be

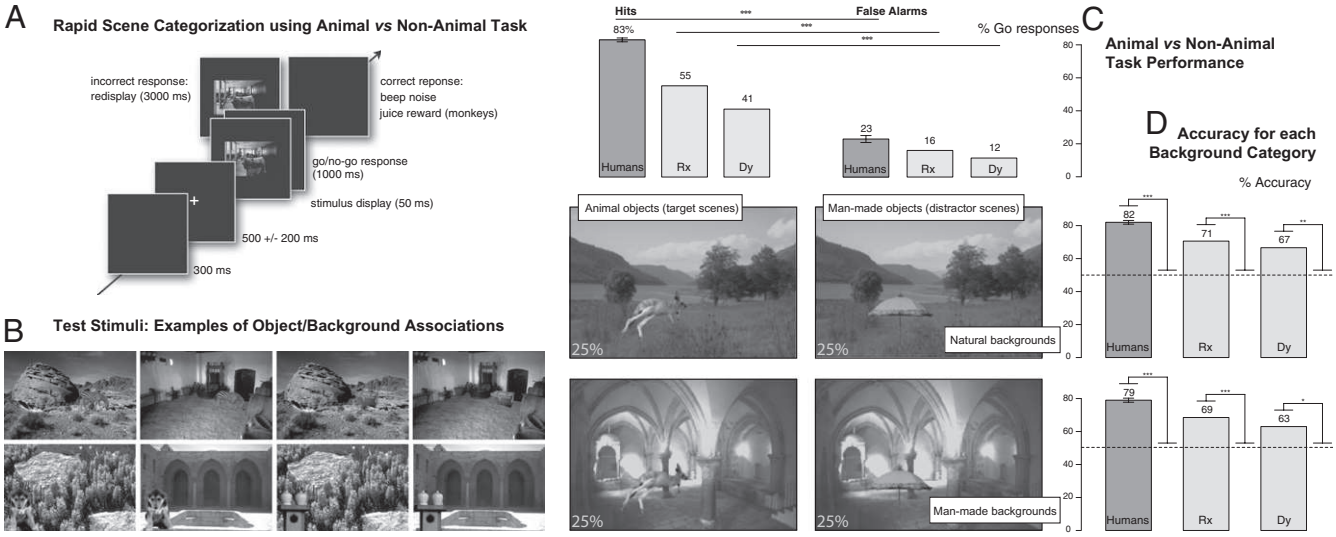

Fig. 1. Animal/Non-Animal categorization performance reached by human and monkey subjects on the first presentation of the test stimuli. (*A*) Animal/Non-Animal task. Stimuli were presented for three frames unmasked on a black screen. When a scene containing an animal was flashed, subjects had to release a button and touch the tactile screen within 1 s (target stimulus, go response); otherwise, they kept their hands on the button (distractor stimulus, no-go response). Correct (go and no-go) responses were rewarded by a noise (and a juice drop for monkeys); incorrect (go and no-go) responses were punished by the redisplay of the incorrectly categorized stimulus for 3 s. (*B*) Two examples of the 192 associations of four stimuli. For each association, the man-made and natural backgrounds had equal average luminance and RMS contrast, and the animal and man-made object vignettes had equal surface, center-of-mass location, luminance, and RMS contrast. (*C*) First-trial hit and false-alarm rates for the group of 11 human subjects and the macaque monkeys Rx and Dy. Bar histograms for hits and false alarms correspond to the stimulus illustrations in the left and right columns, respectively. (Humans: paired *t* test, $t = 26$, df = 10, $P < 10 \times 10^{-9}$; Rx: χ^2 test, $\chi^2 = 122$, $P < 10 \times 10^{-9}$; Dy: χ^2 test, $\chi^2 = 86$, $P < 10 \times 10^{-9}$.) (*D*) First-trial performance accuracy (pooling correct go and no-go responses) computed separately for natural and man-made background stimuli. Bar histograms for the accuracies on natural and man-made backgrounds correspond to the stimulus illustrations in the upper and lower rows, respectively. Dotted line represents chance level; error bars indicate SEM. (χ^2 tests, natural background—humans: contingency table, $\chi^2 = 269$, $n = 11$, $P < 10 \times 10^{-5}$; Rx: $\chi^2 = 13$, $P < 0.0003$; Dy: $\chi^2 = 9$, $P < 0.003$. χ^2 tests, man-made background—humans: contingency table, $\chi^2 = 327$, $n = 11$, $P < 10 \times 10^{-5}$; Rx: $\chi^2 = 11$, $P < 0.001$; Dy: $\chi^2 = 6$, $P < 0.02$.)

rapidly conveyed within the ventral pathway (21) or through the dorsal visual pathway (39) to frontal cortices (42) in order to prime object representations and facilitate subsequent detailed analysis in the ventral visual pathway. Another possibility is that midlevel areas in the ventral visual pathway might use intermediate representations that could be sufficient for categorical judgments (43–45). In both cases, coarse visual information is presumed to be sufficient to trigger prefrontal cortex activity that reflects the precise delineation of categorical boundaries relevant for task performance, as proposed by Freedman, Miller, and collaborators (6, 8).

Object and Context Congruency. Are superordinate visual representations similar in humans and macaques, and do they involve analogous mechanisms? To address these questions, we investigated the interference between object and background categories: we defined as categorically congruent the stimuli that associated congruent superordinate object/context associations (animals pasted on natural backgrounds or man-made objects on man-made backgrounds). Conversely, stimuli that embedded animals in man-made backgrounds or man-made objects in natural backgrounds were considered to be noncongruent (31).

From the very first stimulus presentations, both monkeys and humans exhibited 8% accuracy advantage for congruent compared with noncongruent test stimuli (Table S1). Similar object/context congruence biases have previously been described in humans for rapid scene categorization using color stimuli (27–31); such congruence effects were observed even on the earliest behavioral responses, leading to the suggestion that it could result from feed-forward facilitation between neuronal populations that are usually coactivated because of their selectivity to visual features that are highly likely to co-occur within our environment (27). Following this hypothesis, we predicted that this congruence bias would be robust to short-term practice independently from any improvement in task performance.

Monkey and human performances were thus compared for three consecutive sessions in which all test stimuli where seen only once a day. The monkeys further performed ad libitum for several daily sessions on test stimuli only (9 sessions for Dy and 15 for Rx). As illustrated Fig. 2*A*, we observed a significant accuracy increase with practice in both species, although it was less pronounced in monkey Dy. Such accuracy increase with scene familiarity could be because of a higher success rate on difficult stimuli (34) in particular for incongruent object/background associations. However, no interaction between congruence effect and task practice was observed in either species. In monkeys, a two-way ANOVA (congruence × session) paired by subjects yielded a main effect for categorical congruence [$F(1, 43) = 90$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.17$] and session [$F(10, 43) = 14$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.26$], without interaction between congruence and session [$F(10, 43) = 2$, $P > 0.1$, $\eta^2 = 0.03$]. Similar ANOVA performed over the group of human subjects showed significant effects of categorical congruence [$F(1, 65) = 299$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.39$] and session [$F(2, 65) = 90$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.24$] and no interaction [$F(2, 65) = 3$, $P > 0.05$, $\eta^2 = 0.009$]. Similar results were obtained on reaction times (*SI Materials and Methods*). Thus, although global performance could improve, the impairment observed with incongruent object/background associations was not reduced with practice in either species. This result reinforces the hypothesis of hard-wired (dis)facilitation mechanisms between neurons selective for visual features (not) belonging to the same superordinate categories, as mentioned above.

Are scene statistics the visual features that account for background interaction with object category? We tested this hypothesis using Oliva and Torralba's classification model (24) on the stimuli of the current experiment. This model efficiently implements the major principles of scene gist recognition, considering the global scene features (25) used in a number of recent scene-recognition models (e.g., refs. 46 and 47). First, the simulation was successful in selecting the animal scenes within a set of familiar images that monkeys had categorized over the years (768

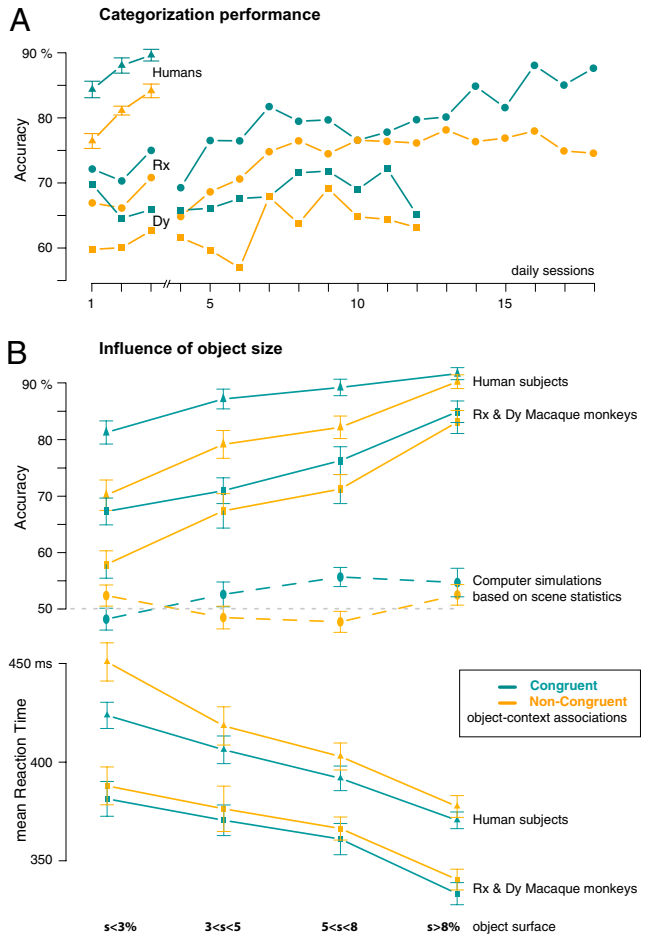

Fig. 2. Global Animal/Non-Animal task performance computed separately for categorical congruent (blue) and noncongruent (orange) object/context associations, reached by human (triangle) and monkey (circle and square) subjects. (*A*) Performance accuracy for consecutive sessions (test stimuli, all trials). Both species exhibited an accuracy increase with practice (linear regressions—humans: coefficient = 3.2, $R^2 = 0.39$, df = 2, $P < 10 \times 10^{-4}$; Rx: coefficient = 0.8, $R^2 = 0.79$, df = 16, $P < 10 \times 10^{-6}$; Dy: coefficient = 0.4, $R^2 = 0.24$, df = 10, $P = 0.06$) but no interaction between categorical congruence and practice (see text). (*B*) Accuracy and mean reaction time as a function of object surface assessed as a percentage of the whole image; stimuli were divided in four sets of equal size. The accuracy performance reached with computer simulations (dashed lines) was computed for the Animal vs. Non-Animal categorization task performed on the test stimulus set, after learning had been achieved by using the monkey's familiar training image set. Dotted line represents chance level; error bars indicate SEM between object/background associations.

photographs, model average accuracy 74%; note, however, that Rx and Dy scored 95% and 96%, respectively, on such familiar images). Second, it was successful in distinguishing between man-made and natural scene backgrounds by using the current 768 test stimuli (averaging 83% accuracy). These results stress the fact that processing image statistics could be a straightforward mechanism for using contextual information. However, the model failed completely at categorizing the manipulated test stimuli as containing animals: it only reached 53% accuracy on congruent and 50% on noncongruent object/context associations (average accuracy vs. chance level, χ^2 test, not significant; congruent vs. noncongruent, χ^2 test, $P < 0.005$).

This failure, however, could be accounted for by the small animal size relative to previous experiments. We tested this hypothesis by calculating simulation results by quartiles of the object sizes, corresponding to 2%, 4%, 7%, and 11% of the whole image.

Although model accuracy increased mildly with object surface, computer simulations only reached 54% accuracy on the largest objects: the model based on scene statistics failed in object categorization.

On the contrary, this analysis performed on subjects' behavioral data showed an important size effect that was similar in both species: performance improved with increasing object size for both accuracy and mean reaction time (Fig. 2*B*; ANOVAs are detailed in *SI Materials and Methods*). Interestingly, the accuracy advantage for congruent compared with noncongruent stimuli was highest for small objects (both species >9%), whereas the smallest accuracy advantage was observed for the biggest objects (both species <2%; Table S2 and ANOVAs in *SI Materials and Methods*). The level of interference between context category and object processing was thus similarly related in both species to the background and object surfaces.

These results suggest that background statistics could play a key role in the context interference observed when humans and monkeys categorize objects in natural scenes. Even if background statistics are not the diagnostic cues used by either species for object categorization, they could play a role analog to the contextual cueing reported by Chun and Jiang in visual search tasks (48): these authors showed that repeated associations between spatial configurations and target locations helped subjects in spatial attention tasks, despite the fact that subjects were not aware of these repeated associations and were not subsequently able to explicitly recognize them. In the present case, similar implicit learning could have occurred through repeated exposures to natural backgrounds associated with the presence of animals. Such implicit learning could thus involve low-level visual cues, including global scene statistics, consistent with their proposed role in triggering the fastest responses during context-categorization tasks (49).

These data suggest a high analogy between monkeys and humans in the way they use visual cues to access categorical representation, the underlying cerebral mechanisms, and their impact on behavior. They suggest the existence of analogous context and object superordinate representations in both species. This proposition and its consequences in terms of processing speed are further tested below.

Fast Access to Abstract Representations and Their Local Interaction. If the processing mechanisms responsible for object/context interactions are analogous in humans and monkeys, any advantage in terms of processing speed should be similar in the two species. More precisely, the underlying hypothesis of feed-forward facilitation implies that this processing speed advantage should be significant from the fastest responses. Indeed, in humans, previous investigations of the influence of object/context congruence on rapid go/no-go scene categorization (27, 31) reported that scenes in which the object conflicted with the surrounding scene backgrounds required an additional processing time of ∼10–20 ms for the earliest responses. We thus took advantage of the severe time constraints imposed by the task to investigate whether the early response onset would be delayed when object and context categories are in conflict. We made the assumption that this delay previously documented in humans would also exist for macaques, despite the facts that the fastest reaction times were associated with the biggest objects, and the biggest objects were associated with the smallest congruence effect on average performance.

For this analysis, a large number of trials was needed that could not be restricted to first-trial performance. For both species, all go responses over all sessions were expressed by using 10-ms time bins. Minimal reaction time (corresponding to the minimal input–output processing time) was then determined as the first time bin for which correct responses significantly exceeded false alarms, using binomial tests. This temporal analysis was performed for each subject individually.

Individual results showed that monkeys Rx and Dy exhibited, respectively, 30-ms and 20-ms delays between minimal reaction time on congruent and noncongruent object/context associations (Table S3 and Fig. S1). In humans, similar results were observed despite a higher variability (0- to 40-ms individual delays), but the reverse effect was never observed; this intrinsic individual variability was likely enhanced by the fact that human subjects performed far fewer trials than the monkeys did. Interestingly, in all subjects, the earliest false alarms were produced when natural backgrounds were presented: scene backgrounds efficiently biased behavior from the fastest responses.

Pooled across subjects (Fig. 3*A*), the distributions of go responses showed a general 50-ms advantage for monkeys compared with humans. However, an equal delay of 30 ms in minimal reaction time was observed in both species between the responses to congruent and noncongruent stimuli (monkey, 200 vs. 230 ms; human, 250 vs. 280 ms): when the object category conflicts with the surrounding context, the additional processing time is similar in the two species from the earliest responses. This finding was further observed with a d' analysis designed to evaluate how accuracy varies with response latency independently of the subject's strategies. Stimulus detectability was computed over time by using cumulative d' scores, by computing $d' = z(\text{hit}) - z(\text{fa})$ by 10-ms time bin (Fig. 3*A Inset*). Data showed that performance with congruent object/context associations reached higher d' values than with noncongruent ones with a similar temporal shift toward shorter response latencies for humans and monkeys. Considered in the frame of signal detection theory, this result indicates that human and macaque cerebral mechanisms are equally sensitive to the congruence of the visual features that determine object and context categories.

To quantify how object size could affect the congruence effect from the fastest responses, we computed minimal reaction times with stimuli containing either the largest or the smallest objects (Fig. 3*B*). Between the earliest responses to small and large objects, a 30-ms delay was observed in both species, showing that "larger is faster" (50) not only on average performance but also for the fastest responses. For these large objects, object/context incongruence delayed the earliest responses by 10 ms in both species (Fig. 3*C*); this value can be compared with the 30-ms average delay for all object sizes in both species reported above.

Such similar delays in monkeys and humans are surprising. Direct behavioral comparisons have, until now, reported macaque response latencies at about two-thirds of the human ones. This principle seems to hold for ocular and manual responses for both absolute latencies as well as for the delays induced by task manipulations (21, 51, 52). A convincing explanation could be the slow speed of intracortical connections that makes brain size a critical factor in determining the time of information transfer between cortical areas (20, 21). If it is the case, the similar delays in humans and monkeys responses reported here are the likely signatures of analogous local computations or, in other words, analogous mechanisms that are embodied in restricted cortical regions all along the processing pathways. For example, the above 30-ms delay related to object size observed during object categorization by human and monkey plausibly reflects similar neural integration mechanisms in the ventral pathways of the two species: indeed Sripati and Olson (53) recently reported that shape selectivity of macaque inferotemporal neurons could develop 30 ms earlier for large stimuli compared with small ones. Here we propose that the highly similar temporal dynamics of object/context interactions observed behaviorally are the signature of analogous fast visual mechanisms that locally process features for object and scene category.

In summary, these results demonstrate first that macaque monkeys can really perform Animal/Non-Animal categorization tasks based on animals as a category and that they can do the task irrespective of scene background content. Moreover, despite their faster reaction times, monkeys exhibited highly similar behavior to human subjects when facing object/context incongruence: there was a similar accuracy impairment in object categorization, similar reaction-time delays observed from the fastest responses, and similar sensitivity to the object/background surface ratio. In both species, the Animal representations used to perform the task

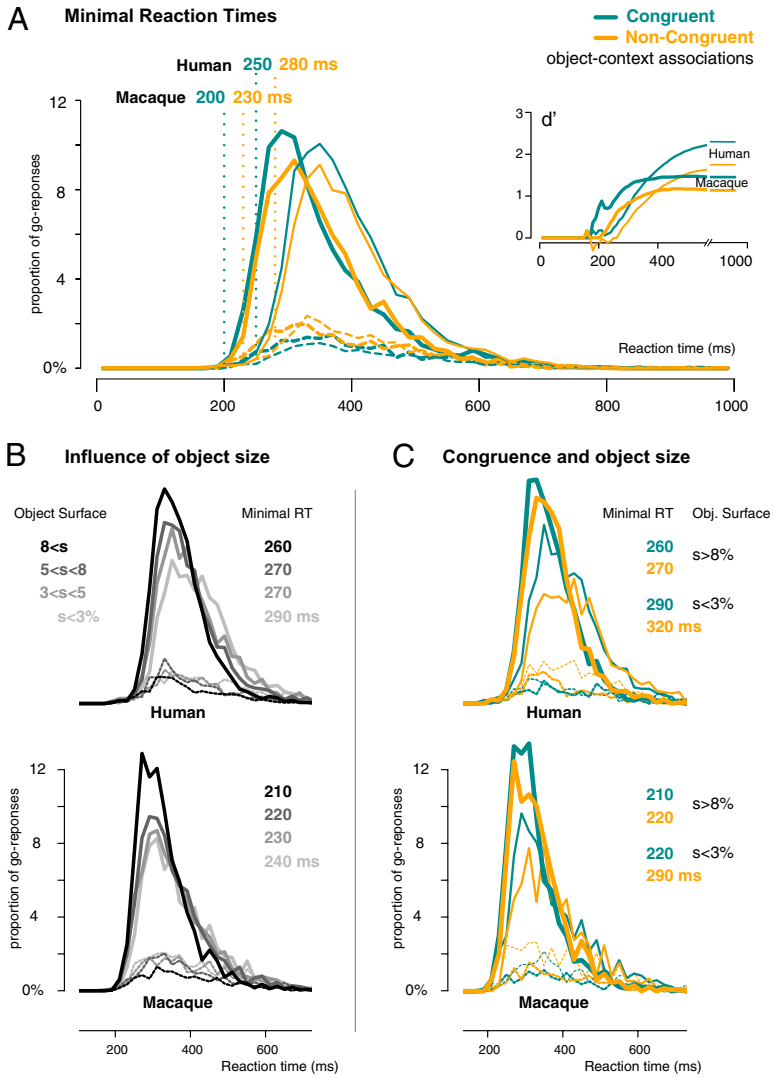

Fig. 3. Reaction-time distributions in the Animal/Non-Animal task computed separately for categorical congruent (blue) and noncongruent (orange) object/context associations. (*A*) Performance speed of humans (thin lines) and macaques (thick lines). Minimal reaction time (vertical dotted lines) was defined as the earliest 10-ms time bin in which hits (continuous lines) significantly outnumber false alarms (dashed lines). (*Inset*) Cumulative d' curves of human and monkey responses to congruent and noncongruent object/context associations. (*B*) Distribution of reaction times as a function of object surface assessed as a percentage of the whole image. For both species, the object surface and associated minimal reaction times are indicated by the gray-level code. (*C*) Distribution of reaction times as a function of object surface (thick lines, large surfaces; thin lines, small surfaces; dashed lines, false alarms) when objects were categorically congruent (blue) or not (orange) with their background context. Associated minimal reaction times are indicated for both species.

are sensitive to background cues but are mostly related to animal feature information that needs to be generalized from a wide range of animal types and shapes. These superordinate representations presumably result from visual neuronal mechanisms that operate in a coarse-to-fine scheme, where scene background and object visual features are locally processed in parallel with early interactions and competition for accumulating evidence in favor of a category that is relevant behaviorally (54).

Nevertheless, it might appear surprising that humans and macaques share these brain mechanisms and early representations. The extreme temporal constraints imposed in the task, in particular the short stimulus duration (50 ms) that restricts the time available for information uptake and the time allowed for response production (1 s), may have emphasized the similarity between humans and monkeys. There is little doubt that, with no time constraints, humans could use more sophisticated strategies that might allow them to perform the task at 100% correct, but when forced to rely on fast early processing of visual information, the cerebral mechanisms used by humans appear to be very similar to those used by monkeys, a "cognitive brick" that would be common to both species. Natural selection may have favored the development of facilitatory mechanisms between populations of object- and feature-selective neurons relatively early in visual processing. If so, one might expect these categorical abilities could be broadly shared across species.

Materials and Methods

Subjects and Initial Training. Two male rhesus monkeys (Rx and Dy, both aged 14 y) performed a go/no-go task to categorize natural scene pictures as containing (or not containing) animals. Initial training followed the procedure reported in Fabre-Thorpe et al. (20): learning was progressive, starting with 10 images and gradually introducing new scenes every day over a period of several weeks until both monkeys were performing well on ∼300 stimuli. Although the monkeys' motivation and level of reward were kept stable by randomly interleaving familiar and new stimuli, the recurrent introduction of new stimuli (usually 10–20%) forced the monkeys to look for an underlying rule to produce the adequate response rather than to rely on stimulus memorization. Both monkeys had been trained for intermittent periods on the Animal/Non-Animal task since 2005, but monkey Rx had first been trained on a Food/Non-Food task since 1996 (19–21, 55). For both monkeys, the set of familiar stimuli included at least 750 stimuli. All procedures conformed to French and European standards concerning the use of experimental animals; protocols were approved by the regional ethical committee for experimentation on animals (agreement ref. MP/05/05/01/05).

Eleven human subjects (aged 23–50 y, 37 y average, five females) performed the same categorization task using the same experimental setup. Stimulus size in pixels and display were identical, as were behavioral control (button release, screen touch, and the 3-s stimulus display that followed incorrect decisions). Correct decisions were indicated by a beeping noise only.

Stimuli. In familiar and test stimuli, the Animal and Man-Made Object superordinate categories include very varied exemplars. The Animal category included mammals (57% and 65% in familiar and test stimuli, respectively),

birds (23% and 15%), insects (2% and 6%), reptiles and amphibians (8%), and fish and crustaceans (10% and 6%). In the test stimuli, man-made objects included means of transport (16%), urban furniture (15%), house furniture (16%), kitchen and water containers (17%), tools and toys (13%), interior decoration (8%), and other various objects (15%). The familiar stimuli were commercial photographs in which animals were usually large (or even very large), centered, and focused by the photographer. In the test stimuli, a wide range of object sizes and locations were used. Views of all of the test stimuli with associated first-trial performances and a subset of familiar stimuli are available at http://cerco.ups-tlse.fr/~denis/fizePNAS2011. For further details about test stimulus generation, see *SI Materials and Methods*.

Procedure. In a session, human subjects performed the Animal/Non-Animal categorization task once using the 768 test stimuli presented randomly (about half an hour). Each human subject performed three sessions on a daily basis to assess the robustness of the behavioral measures with three repetitions. For monkeys, test stimuli were introduced progressively, intermixed with familiar stimuli. Twelve daily sessions were needed to record monkey performance on the complete test stimulus set presented three times. Monkeys were further tested on daily sessions (9 for Dy and 15 for Rx) using the 768 test stimuli randomly presented ad libitum. Monkeys Dy and Rx performed, respectively, a total of 7,940 and 14,560 trials on test stimuli.

Computer Simulations. Simulations used the code distributed by Oliva and Torralba (24) using the default software parameters. Each simulated task included 500 simulations: the performance accuracy indicated in the text result from their average value (SEM ranged from 1.07% to 1.2%). Each simulation involved randomly shuffling the stimuli into two equal sets that were considered for the subsequent phases of learning and testing.

More detailed descriptions of the methods can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Simon Thorpe, Rufin VanRullen, Leila Reddy, and Thomas Serre for valuable comments on the manuscript. Computer simulations were possible thanks to the code freely provided by Aude Oliva and Antonio Torralba. This work was supported by the Centre National de la Recherche Scientifique; the Université de Toulouse, Université Paul Sabatier; and the Fondation pour la Recherche Médicale. The Délegation Générale de l'Armement provided financial support to M.C.

1. Bovet D, Vauclair J (2001) Judgment of conceptual identity in monkeys. *Psychon Bull Rev* 8:470–475.
2. Herrnstein RJ, Loveland DH (1964) Complex visual concept in the pigeon. *Science* 146: 549–551.
3. Zayan R, Vauclair J (1998) Categories as paradigms for comparative cognition. *Behav Processes* 42:87–99.
4. Lazareva OF, Wasserman EA (2008) *Learning Theory and Behavior*, ed Menzel R (Elsevier, Oxford), pp 197–226.
5. Minamimoto T, Saunders RC, Richmond BJ (2010) Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron* 66:501–507.
6. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312–316.
7. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2002) Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *J Neurophysiol* 88:929–941.
8. Cromer JA, Roy JE, Miller EK (2010) Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66:796–807.
9. Vogels R (1999) Categorization of complex visual images by rhesus monkeys. Part 1: Behavioural study. *Eur J Neurosci* 11:1223–1238.
10. Schrier AM, Angarella R, Povar ML (1984) Studies of concept formation by stumptailed monkeys: Concepts humans, monkeys, and letter A. *J Exp Psychol Anim Behav Process* 10:564–584.
11. Vauclair J, Fagot J (1996) Categorization of alphanumeric characters by baboons (*Papio papio*): Within and between class stimulus discrimination. *Curr Psychol Cogn* 15:449–462.
12. Sigala N, Gabbiani F, Logothetis NK (2002) Visual categorization and object representation in monkeys and humans. *J Cogn Neurosci* 14:187–198.
13. Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563.
14. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139.
15. Weiskrantz L, Saunders RC (1984) Impairments of visual object transforms in monkeys. *Brain* 107:1033–1072.
16. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23: 5235–5246.
17. 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.
18. Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–320.
19. 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 Res* 40: 2187–2200.
20. Fabre-Thorpe M, Richard G, Thorpe SJ (1998) Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9:303–308.
21. Macé MJ, Richard G, Delorme A, Fabre-Thorpe M (2005) Rapid categorization of natural scenes in monkeys: Target predictability and processing speed. *Neuroreport* 16:349–354.
22. Roberts WA, Mazmanian DS (1988) Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J Exp Psychol Anim Behav Process* 14:247–260.
23. Fabre-Thorpe M (2003) Visual categorization: Accessing abstraction in non-human primates. *Philos Trans R Soc Lond B Biol Sci* 358:1215–1223.
24. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.
25. Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. *Prog Brain Res* 155:23–36.
26. Hughes HC, Nozawa G, Kitterle F (1996) Global precedence, spatial frequency channels, and the statistics of natural images. *J Cogn Neurosci* 8:197–230.
27. Joubert OR, Fize D, Rousselet GA, Fabre-Thorpe M (2008) Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J Vis* 8:11.
28. Davenport JL, Potter MC (2004) Scene consistency in object and background perception. *Psychol Sci* 15:559–564.
29. Davenport JL (2007) Consistency effects between objects in scenes. *Mem Cognit* 35: 393–401.
30. Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? *J Vis* 7:10.
31. Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M (2007) Processing scene context: Fast categorization and object interference. *Vision Res* 47:3286–3297.
32. Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
33. Minamimoto T, La Camera G, Richmond BJ (2009) Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. *J Neurophysiol* 101:437–447.
34. 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. *J Cogn Neurosci* 13:171–180.
35. Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD (2003) Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 7:84–91.
36. Humphreys GW, Forde EM (2001) Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behav Brain Sci* 24: 453–476, discussion 476–509.
37. Macé MJ, Joubert OR, Nespoulous JL, Fabre-Thorpe M (2009) The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE* 4:e5927.
38. Fabre-Thorpe M, *The Making of Human Concepts*, eds Mareschal D, Quin PC, Lea S (Oxford Univ Press, Oxford).
39. Bullier J (2001) Integrated model of visual processing. *Brain Res Brain Res Rev* 36: 96–107.
40. Schyns PG, Oliva A (1994) From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychol Sci* 5:195–200.
41. Navon D (1977) Forest before trees: The precedence of global features in visual perception. *Cognit Psychol* 9:353–383.
42. Bar M, et al. (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103:449–454.
43. Mirabella G, et al. (2007) Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron* 54:303–318.
44. Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. *Nat Neurosci* 5:682–687.
45. Delorme A, Richard G, Fabre-Thorpe M (2010) Key visual features for rapid categorization of animals in natural scenes. *Front Psychol* 1:21.
46. Grossberg S, Huang TR (2009) ARTSCENE: A neural system for natural scene classification. *J Vis* 9:6.
47. Siagian C, Itti L (2007) Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Trans Pattern Anal Mach Intell* 29:300–312.
48. Chun MM, Jiang Y (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognit Psychol* 36:28–71.
49. Joubert OR, Rousselet GA, Fabre-Thorpe M, Fize D (2009) Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *J Vis* 9:2.
50. Vogels R (2009) Visual perception: Larger is faster. *Curr Biol* 19:R691–R693.
51. Busettini C, Fitzgibbon EJ, Miles FA (2001) Short-latency disparity vergence in humans. *J Neurophysiol* 85:1129–1152.
52. Fischer B, Weber H (1993) Express saccades and visual attention. *Behav Brain Sci* 16: 553–567.
53. Sripati AP, Olson CR (2009) Representing the forest before the trees: A global advantage effect in monkey inferotemporal cortex. *J Neurosci* 29:7788–7796.
54. Perrett DI, Oram MW, Ashbridge E (1998) Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition* 67:111–145.
55. Macé 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.

Supporting Information

Fize et al. 10.1073/pnas.1016213108

SI Materials and Methods

Test Stimuli. Test stimuli were 768 gray-level scenes (576 × 384 pixels) that had the cut out of either an animal or a man-made object pasted onto either a natural or a man-made scene background. Part of the 384 man-made object vignettes had been collected and used by Joubert et al. (1) and were all extracted from either the Corel Stock Photo or Hemera Photo Objects libraries. Animal targets included fish, birds, mammals, reptiles, insects, and amphibians, whereas the man-made objects distracters included means of transport, city and house furniture, as well as various objects such as tools or monuments. The 384 picture backgrounds were mainly selected from the pool used by Joubert et al., with additional scenes selected because of their angle of view. Natural scene backgrounds included pictures from sea, sky, mountains, and forest, and landscape views from various geographical regions; man-made scene backgrounds included inside and outside urban views, such as building and street pictures, airports, or harbors.

The object vignettes and scene backgrounds were organized in 192 object and scene associations of four test stimuli (Fig. 1*B*). Each association consisted of an animal and a man-made object vignette selected for their similar real size or distance from the photographer, and a natural and a man-made scene background selected to have a similar depth of field. Each object vignette was pasted on both scene backgrounds, providing a set of four test stimuli. Thus, each of the 192 associations provided two object/context congruent stimuli (an animal pasted on a natural background, a man-made object on a man-made background) and two object/context noncongruent stimuli (the animal on the man-made background, the man-made object on the natural background).

Each four test stimuli corresponding to an association were generated simultaneously with a homemade image-processing software (written using Matlab). Particularly, within each four test stimulus associations, the natural and man-made scene backgrounds were adjusted to equalize average luminance and RMS contrast. The associated animal and man-made object vignettes were adjusted to equalize surface areas (alpha layer surfaces in pixels), locations (equal center-of-mass coordinates), average luminance, and RMS contrast. The luminance and contrast values were adjusted to the average values of the scene background surfaces that were occluded by the vignettes, so that the vignettes were fully and equally integrated to their scene backgrounds, with the goal of equalizing saliency. The 768 test stimuli were thus equalized pairwise for foreground objects and background scenes, as well as low-level statistics.

Procedure. Monkey subjects were placed in a semidark room, restrained in a primate chair (Crist Instruments) but with their heads free, and sat ~30 cm away from a 1,024 × 768 tactile screen (driven by the programmable VSG 2 graphics board; Cambridge Research Systems) controlled by a PC-compatible computer. The subjects placed one hand on a response pad located below the screen at waist level to start stimulus presentation. Pictures were then flashed centrally for 50 ms (three frames at 60 Hz, noninterlaced; the stimuli for monkeys were covering ~42 × 28 degrees of visual angle) on a black background, with a 1.5- to 2-s random intertrial interval between successive images. Because of human arm length and requested movement for behavioral response, the screen on which the same stimuli were displayed was placed at 50 cm away from human subjects (at this distance, stimuli covered 25 × 17 degrees of visual angle). These brief presentations prevented exploratory eye movements and constrained the time available for information uptake. Subjects had 1 s to release the response pad and touch the tactile screen when they detected an animal in the flashed image (target stimuli, go response), otherwise they had to keep pressing the button (distractor stimuli, no-go response). Any response after 1 s was considered as a no-go response. A drop of fruit juice for monkeys and a beeping noise rewarded correct go- or no-go decisions. Incorrect go and no-go decisions were followed by a 3-s display of the incorrectly classified stimulus, delaying the next trial and the next possibility of reward and allowing time for ocular exploration. Reaction times (delay between stimulus onset and response pad release) and accuracy were recorded online; stimulus presentation and behavioral monitoring used custom software. Monkeys performed for as long as they wanted (typically 1 h for 1,000 trials) during a daily session.

For monkey subjects, test stimuli were introduced progressively, intermixed with a set of familiar stimuli used for training. The first 12 daily sessions were designed to this end: 100 familiar stimuli (from the pool of pictures categorized many times by the monkeys) were first used to stabilize monkeys' performance; then, 192 test stimuli were introduced randomly together with 384 familiar stimuli (2/3 in proportion); the last part of a session used more than 700 familiar stimuli so that monkeys could perform as long as they wanted until satiation. Stimulus randomization was recomputed for each session. The same 192 test stimuli (one-fourth of the total test stimulus set) were used for three successive sessions, so that, as for human subjects, the robustness of any behavioral effect could be assessed on three successive presentations. Twelve sessions were thus needed to record monkey performance on the complete test stimulus set presented three times. For clarity in the results and analyses, we pooled together the trials performed on the entire test stimulus set (for a given ordinal presentation).

Thereafter, the monkeys continued performing for a number of daily sessions (9 for Dy and 15 for Rx); these daily sessions only used the 768 test stimuli randomly presented ad libitum. On the test stimulus set, monkeys Dy and Rx performed a total of 7,940 trials and 14,560 trials, respectively.

Data Analysis. For each subject, trials with reaction times longer than 3 SDs were not considered for reaction-time analysis (~0.1%). Validation of all ANOVAs was performed by using residual analysis (variance homogeneity and normality); reaction times were log-transformed. Global results of Fig. 2 were assessed using paired *t* tests, contingency tables, and ANOVAs. Global accuracy and mean reaction times as a function of object size (Fig. 3*B*) were first computed per image considering all subjects' trials; mean and SEM were further computed for each size (object surfaces grouped per quartiles). Minimal reaction times were computed by using binomial tests considering the cumulated average theoretical proportion of response for each 10-ms time bin.

Effect of practice and object/background category congruence on response speed (Fig. 2A). On reaction times that correspond to human and monkey performances illustrated in Fig. 2*A*, monkeys exhibited a small categorical congruence effect (two-way ANOVA, congruence × session) [$F(1, 43) = 9$, $P < 0.02$, $\eta^2 = 0.01$] and session effect [$F(10, 43) = 26$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.28$], without interaction between congruence and session [$F(10, 43) = 1$, $P > 0.5$, $\eta^2 = 0.01$]. Similar ANOVA on mean reaction times performed over the group of human subjects showed significant but small effects of categorical congruence [$F(1, 65) = 46$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.01$] and session [$F(2, 65) = 25$, $P < 10 \times 10^{-4}$,

$\eta^2 = 0.01$] and no interaction [$F(2, 65) = 3$, $P > 0.05$, $\eta^2 = 0.002$]. Thus, for reaction time, as for accuracy, the performance impairment observed when object and background belonged to different man-made and natural categories was not reduced with practice in either species.

Effect of object surface on categorization accuracy (Fig. 2B). Two-way ANOVAs (object surface × categorical congruence) yielded a strong main effect of object surface on performance accuracy in both species [humans: $F(3, 87) = 231$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.54$; monkeys: $F(3, 15) = 1{,}847$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.66$] and confirmed the main effect of categorical congruence [humans: $F(1, 87) = 267$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.21$; monkeys: $F(1, 15) = 508$, $P < 10 \times 10^{-3}$, $\eta^2 = 0.06$]. The interaction between object surface and congruence was significant in both species [humans: $F(3, 87) = 23$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.05$; monkeys: $F(3, 15) = 55$, $P < 0.005$, $\eta^2 = 0.02$]; the largest accuracy advantage for congruent compared with noncongruent stimuli was recorded for the smallest objects (both species >9%) compared with the biggest objects (both species <2%). Table S2 shows data and significance of the effect using paired *t* tests.

Effect of object surface on reaction times (Fig. 2B). As for accuracy, two-way ANOVAs (object surface × categorical congruence) on reaction times showed similar trends. In both species, there were main effects of object surface [humans: $F(3, 87) = 344$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.22$; monkeys: $F(3, 15) = 13$, $P < 0.004$, $\eta^2 = 0.18$]. In humans, categorical congruence affected reaction times [$F(1, 87) = 79$, $P < 10 \times 10^{-4}$, $\eta^2 = 0.02$] with a small interaction with object surface [$F(3, 87) = 6$, $P < 0.003$, $\eta^2 = 0.003$]. No effect of categorical congruence nor any interaction with object surface were observed on response speed in monkeys [main: $F(1, 15) = 0.85$, $P > 0.5$, $\eta^2 = 0.003$; interaction: $F(3, 15) = 0.25$, $P > 0.8$, $\eta^2 = 0.004$].

Computer Simulations. Simulations used the code distributed by A. Oliva and A. Torralba (2), which is available at http://people.csail.mit.edu/torralba/code/spatialenvelope/. All simulated task performances used the default software parameters. Each simulated task included 500 simulations: the performance accuracy indicated in the text result from their average value (SEM ranged from 1.07% to 1.2%). Each simulation involved randomly shuffling the stimuli into two equal sets that were considered for the subsequent phases of learning and testing. The 768 familiar stimuli used for the monkey experiments (Corel Stock Photo) were used for a first simulation of the Animal vs. Non-Animal categorization task; the 768 test stimuli were use to simulate a natural vs. man-made scene context-categorization task. A simulated Animal vs. Non-Animal categorization task with the test stimuli was then used to simulate monkey's performance on the first presentations of the test stimuli. Thus, for each simulation, 384 stimuli were randomly selected within the pool of 768 familiar stimuli and used in the learning phase; 384 stimuli were randomly selected within the pool of the 768 test stimuli and used in the testing phase. We also checked that the software obviously failed to learn the Animal vs. Non-Animal task using the test stimuli only.

1. Joubert OR, Fize D, Rousselet GA, Fabre-Thorpe M (2008) Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J Vis* 8:11.

2. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.

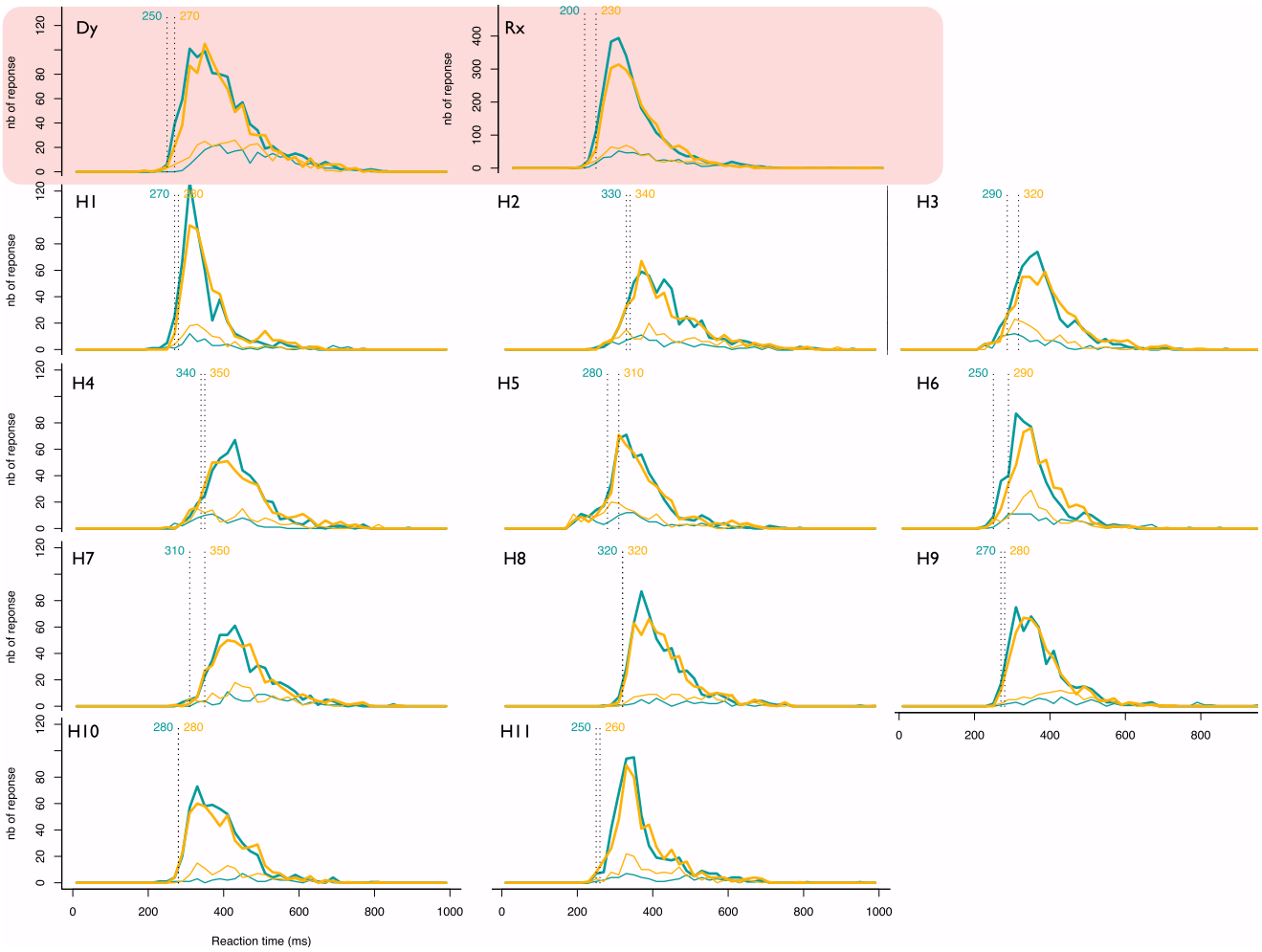

Fig. S1. Individual histogram distributions of reaction times (all trials). Number of hits (thick lines) and false alarms (thin lines) as a function of response latency (20-ms time bin). Dotted lines indicate minimal reaction times (time at which correct go responses significantly outnumber false alarms). Minimal reaction times are shorter for congruent (blue) than noncongruent (orange) object/context associations. Dy and Rx are macaque monkeys, and H1–11 are human subjects.

Table S1. Individual results using first-trial performance on the test stimulus set

	Accuracy, %			Median reaction time, ms			Mean reaction time, ms		
Subjects	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference
Dy	69.8	59.8	10	386	403	−18	411	421	−10
Rx	72	66.8	5.2	324	367	−43	354	378	−25
H1	87.1	81.6	5.5	325	343	−19	343	362	−19
H2	78.9	70.4	8.5	411	412	−1	432	433	−1
H3	84.2	77.7	6.5	377	391	−14	392	396	−4
H4	80.2	75.1	5.1	405	403	2	422	424	−2
H5	80.8	73.9	6.9	366	376	−11	372	382	−10
H6	81.5	73.2	8.3	317	341	−24	332	351	−19
H7	81.1	74.2	6.9	469	453	16	480	471	9
H8	90.3	81.8	8.5	409	411	−2	425	431	−6
H9	88.5	77.3	11.2	352	349	4	363	361	2
H10	89.3	81.1	8.2	428	444	−16	443	448	−4
H11	86.6	74	12.6	330	330	0	348	354	−6
Mean									
Monkeys	70.9	63.3	7.6	355	385	−30	382	400	−17
Humans	84.4	76.4	8	381	386	−6	396	401	−6
SE									
Monkeys	1.1	3.5	2.4	31	18	13	29	21	8
Humans	1.2	1.2	0.7	15	13	4	14	13	2

Dy and Rx are macaque monkeys, and H1–11 are human subjects.

Table S2. Global accuracy for each individual (all trials) as a function of object size

	Object size congruence			
Subjects	$s < 3\%$ congruent	$s < 3\%$ noncongruent	$s > 8\%$ congruent	$s > 8\%$ noncongruent
Humans				
H1	85.8	74.5	94.4	94.4
H2	81.3	66.6	88.9	87.0
H3	84.5	70.0	89.2	88.4
H4	75.8	66.9	89.2	90.6
H5	78.6	69.6	85.8	86.4
H6	78.1	71.7	89.0	89.6
H7	74.9	67.9	89.2	87.1
H8	86.6	69.5	96.1	94.4
H9	77.6	70.6	95.1	89.6
H10	86.7	74.9	96.5	93.8
H11	82.5	68.8	95.1	91.9
Mean	81.1%	70.1%	91.7%	90.3%
SE	1.3%	0.8%	1.1%	0.9%
Congruent−noncongruent	**11.0%**		**1.4%**	
Paired t	$P < 10 \times 10^{-6}$		$P < 0.022$	
Monkeys				
Dy	60.5	51.9	79.7	78.3
Rx	70.9	61.3	87.5	85.7
Mean	65.7%	56.6%	83.6%	82.0%
SE	5.2%	4.7%	3.9%	3.7%
Congruent−noncongruent	**9.1%**		**1.6%**	
Paired t	$P < 0.018$		$P < 0.031$	

Object size is indicated as a percentage of stimulus surface. Accuracy scores are significantly higher when object and context belong to congruent superordinate categories, but this congruence effect is much larger for small objects. Bold indicates the size of the congruence effect on accuracy scores. Dy and Rx are macaque monkeys, and H1–11 are human subjects.

Table S3. Individual results (all trials)

	Accuracy, %			Median reaction time, ms			Mean reaction time, ms			Minimal reaction time, ms		
Subjects	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference
Dy	68.2	63.5	4.6	379	381	-2	400	403	-3	250	270	-20
Rx	78.4	73.2	5.2	306	317	-11	327	333	-6	200	230	-30
H1	89.9	84.6	5.2	326	340	-14	344	362	-18	270	280	-10
H2	85.4	77.5	7.9	413	410	3	430	434	-4	330	340	-10
H3	87.1	80.8	6.3	365	383	-18	378	398	-20	290	320	-30
H4	84.5	79.4	5.1	427	427	0	437	445	-8	340	350	-10
H5	83.9	77.9	6.0	348	352	-4	359	368	-9	280	310	-30
H6	85.0	78.6	6.4	340	357	-17	353	369	-16	250	290	-40
H7	82.9	77.1	5.8	435	443	-8	451	459	-8	310	350	-40
H8	91.6	83.5	8.1	400	411	-11	422	430	-8	320	320	0
H9	88.6	80.1	7.7	355	360	-5	371	375	-4	270	280	-10
H10	92.5	84.2	8.2	377	381	-4	389	395	-6	280	280	0
H11	89.5	80.9	8.6	348	354	-6	371	377	-6	250	260	-10
Mean												
Monkeys	73.3	68.3	4.9	343	349	-7	364	368	-5	225	250	-25
Humans	87.3	80.4	6.8	376	383	-8	391	401	-10	290	307	-17
SE												
Monkeys	5.1	4.9	0.3	37	32	5	37	35	2	25	20	5
Humans	0.9	0.8	0.4	11	10	2	11	11	2	9	9	4

Average performance for each subject on the test stimulus set (see text and Fig. S1). Dy and Rx are macaque monkeys, and H1–11 are human subjects.

2.3 Discussion

Les résultats décrits dans cet article peuvent être interprétés à deux niveaux.
Au niveau expérimental le plus concret, on aborde le problème de l'utilisation de photographies pour étudier les capacités de catégorisation naturelle chez l'animal. On montre ainsi que dans cette tâche de catégorisation rapide, les macaques rhésus réalisent bien une catégorisation superordonnée haut niveau, en se basant sur des traits diagnostiques des objets et non sur des statistiques d'images bas niveau co-occurrentes.

A un niveau théorique plus général, on cherche à savoir si deux espèces phylogénétiquement proches peuvent partager des représentations mentales à l'origine de fonctions cognitives similaires. En engageant deux espèces de primates dans un protocole cognitif strictement identique et en rapportant systématiquement des résultats similaires entre les deux espèces au fur et à mesure des analyses, cette étude argumente fortement en faveur de l'existence de représentations visuelles communes à l'homme et au singe. Des représentations relativement grossières, où la nature du contexte peut influencer la catégorie de l'objet, mais qui sont suffisantes pour détecter rapidement la présence d'un animal dans une scène complexe.

2.3.1 Catégorisation rapide de scène naturelles : une « vraie» catégorisation?

Nous avons parlé dans l'introduction de ce chapitre des critères nécessaires pour montrer une « vraie » catégorisation. Le premier est celui de généralisation à de nouveaux exemplaires. La généralisation est vérifiée dans cette étude en montrant que les performances des singes sont au-dessus de la chance pour catégoriser les nouveaux stimuli dès la première présentation.

Le second critère, qui n'est pas rapporté explicitement dans l'article, est celui de discrimination intra-classe et notamment de discrimination intra-classe entre stimuli d'entraînement et de test (Vauclair et Fagot 1996). Il est essentiel pour éliminer l'hypothèse d'une simple discrimination dans laquelle le singe confondrait tous les stimuli d'une même catégorie. Il y a plusieurs arguments en faveur de cette discrimination intra-classe. D'abord la baisse considérable (environ

20%) de la précision sur le set de test par rapport au set d'entraînement, indique que les stimuli manipulés sont plus difficiles à catégoriser pour les singes et par conséquent relativement différents. Ensuite, nous sommes ici dans une catégorisation de type superordonné, et dans ce type de catégorisation les stimuli d'une même classe sont par définition d'une grande variabilité. Il y a peu de ressemblance physique entre un banc de poisson et un gros plan de la tête d'un chien. Mais est-ce que les singes perçoivent bien ces différences ? Une étude non publiée de Macé et Fabre-Thorpe le confirme. Effectivement, un des deux singes utilisés pour notre article a été entraîné à réaliser une catégorisation basique dans laquelle il devait répondre spécifiquement sur des oiseaux avec comme distracteur des animaux non-oiseaux. Le singe a parfaitement réalisé cette tâche en atteignant une précision de 84%. Il était donc capable de discriminer les oiseaux des autres stimuli de la classe animale.

Malgré les temps de réaction très courts avec lesquels les singes réalisent la tâche de catégorisation rapide, cette tâche remplit tous les critères nécessaires pour parler de véritable catégorisation. La catégorisation rapide nous offre donc un moyen idéal pour étudier les mécanismes suffisants à la formation de concepts visuels.

2.3.2 Pourquoi les singes peuvent-ils être influencés par une association object/contexte ?

Un des principes de la tâche de catégorisation rapide est d'utiliser des images naturelles. Jusqu'à récemment les nombreuses études utilisant ce paradigme ont porté sur la catégorisation d'objets (animaux, oiseaux, véhicules, visages...). Or dans des photographies de scènes naturelles, ces objets apparaissent toujours dans un contexte (une vache dans un pré, une voiture dans la rue, un oiseau dans le ciel...). Il paraît donc légitime de se poser la question de l'influence du contexte sur la catégorisation d'objets. En 2007, Joubert, Rousselet, Fize et Fabre-Thorpe montrent, chez l'homme, que deux grandes catégories de contextes (naturel et artificiel), peuvent être catégorisées avec des temps de réaction aussi rapides (390 ms) que ceux nécessaires à la catégorisation d'objets (400 ms). Ces deux grands domaines d'informations (objets et scènes) sont donc traités simultanément et par conséquent peuvent potentiellement s'influencer.

Cette hypothèse a été vérifiée peu de temps après par la même équipe. Chez l'homme, la catégorisation rapide animal/non animal est réalisée avec des performances inférieures, (précision plus faible et temps de réaction plus longs) lorsque les stimuli présentent une incongruence entre l'objet et le contexte (un animal sur un fond artificiel ou un objet sur un fond naturel) que dans le cas où ils présentent une congruence objet/contexte (un animal sur un fond naturel ou un objet sur un fond artificiel) (Joubert et al. 2008).

D'autres espèces de primates sont également sensibles au contexte dans lequel se trouve l'objet. Martin-Malivel et Fagot (2001) montrent l'existence de concepts multimodaux "homme" et "babouins" chez le babouin. Quatre babouins sont engagés dans cette expérience et deux seulement montrent un effet d'amorçage réellement lié au concept de babouin ou d'homme. En effet, les auteurs démontrent que les réponses des deux autres animaux utilisaient d'autres caractéristiques des stimuli: l'un d'eux utilisait la couleur (plus d'amorçage si les amorces étaient présentées en noir et blanc) et le second utilisait le contexte. Pour ce dernier animal, l'amorçage disparaissait lorsque l'amorce ne présentait plus l'image dans sa totalité mais seulement les sujets 'hommes" ou "babouins" détourés (sans contexte) (Martin-Malivel thèse 2000). Les singes travaillant avec des images naturelles peuvent donc résoudre leur tâche en basant leurs réponses sur des éléments différents de l'image.

Même chez l'homme, le système visuel est capable d'extraire des régularités statistiques (invariances) ou des co-occurrences d'objets pour prédire l'apparition d'autres objets (Fiser et Aslin 2001). Ces covariations peuvent êtres apprises inconsciemment puis utilisées de manière implicite pour guider certains comportements notamment dans des tâches de recherche visuelle (Chun et Jiang 1999).

Enfin, l'influence du contexte est aussi retrouvée au niveau neuronal. Dès le cortex visuel primaire (V1), le taux de décharge de certains neurones varie en fonction de la luminance présente hors de leur champ récepteur (Paradiso et al. 2006).

Dans le cortex inféro-temporal (IT), Murczek et Sheinbeirg (2007) ont montré qu'au niveau du codage d'objets ce phénomène existe aussi. La réponse d'un neurone sélectif à un objet particulier est amplifiée lorsque ce dernier est présenté parmi d'autres objets familiers. Au contraire, si ce même objet est entouré d'objets complètement nouveaux pour le singe, la réponse du neurone est plus faible.

2.3.3 Qu'est ce que les animaux comprennent dans une image ?

Les photographies sont d'usage courant pour l'étude des capacités de catégorisation ou de reconnaissance d'objets naturels chez l'animal. Néanmoins, si nous avons appris qu'une photo, un dessin ou une peinture sont des représentations 2D du monde réel, ce n'est pas le cas des animaux, et on est en droit de se demander quel est leur niveau de compréhension de ces images. Fagot et al. (2009) définissent 3 niveaux :

- Le *niveau d'indépendance* dans lequel l'animal ne fait aucune association entre la photographie et l'objet qu'elle représente. Dans ce cas, il traite le stimulus comme un pattern visuel abstrait dans lequel il doit extraire une caractéristique lui permettant de faire la tâche. Par exemple, D'Amato et Van Sant (1988) ont montré que dans une tâche de catégorisation de stimuli contenant des êtres humains, les capucins peuvent répondre lorsque l'image contient une tâche rougeâtre.
- Le *niveau de confusion*. L'animal ne fait pas de distinction entre l'image et ce qu'elle représente. Ainsi les babouins et même les gorilles à qui l'on présente pour la première fois une image de banane, s'en saisissent et la mettent à la bouche, comme ils feraient avec une vraie banane pour la manger (Parron et al. 2008).
- Le *niveau d'équivalence*. L'animal associe l'image à l'objet qu'il représente. Ce niveau qui nous paraît le plus évident, n'est pas forcément le plus facile à démontrer. Une étude récente apporte des preuves assez convaincantes d'un traitement d'équivalence entre objets et images chez le capucin (Truppa et al. 2009). En utilisant un protocole d'appariement différé, ils montrent que les capucins réussissent parfaitement l'association entre l'objet réel et sa représentation par une image (et vice versa), même lorsque l'image n'est plus qu'un dessin du contour de l'objet.

Quel est le niveau de compréhension des images auquel nos singes rhésus accèdent dans la tâche de catégorisation rapide? Qu'est-ce que représente pour eux une girafe dans la savane, un oiseau dans le ciel ou un banc de poissons devant un récif de corail ? Mis à part leurs congénères dans les enclos et quelques humains, nos singes n'ont absolument aucune expérience avec d'autres types d'animaux. La seule *expérience indirecte* qu'ils aient pu avoir avec la diversité du règne animal est celle que des films animaliers, diffusés sur un écran de télévision dans l'animalerie, auraient pu leur donner.

Une réponse facile aux questions que l'on se pose ici est donc : puisqu'ils n'ont pas d'*expérience directe* avec les objets que les images représentent, ils ne peuvent être ni dans un mode de confusion ni dans un mode d'équivalence mais seulement dans un mode d'indépendance. Comme dans l'étude de D'Amato et Van Sant (1988) nos singes auraient appris un pattern visuel caractéristique des images d'animaux et l'utiliseraient pour réaliser la tâche. Oui mais lequel ? Ce n'est pas la couleur (Delorme et al. 2001) ni les statistiques globales de l'image ou l'information contextuelle comme l'attestent les résultats présentés ici. L'étau se resserre. D'autant plus que nous rapportons aussi ici que les singes se focalisent vraiment sur l'animal en lui-même pour répondre, et que les représentations qu'ils mettent en jeu sont extrêmement similaires à celles que les sujets humains utilisent. Est-ce que, dans la tâche de catégorisation rapide, les hommes ne se serviraient plus de leurs connaissances d'équivalence entre l'image et l'objet réel animal mais, comme les singes, utiliseraient un simple pattern visuel indépendant de l'objet qu'il représente ? C'est une explication possible. Une autre explication, pas forcément exclusive, est que connaissance ne rime pas forcément avec expérience.

Effectivement, l'évolution peut avoir façonné le cerveau pour le doter de quelques facultés « innées » indispensables à la survie. Parmi ces facultés, la détection rapide des animaux figure comme un candidat potentiel. C'est ce que suggère fortement une étude récente chez l'homme (New et al. 2007). En utilisant un protocole de détection de changement, dans lequel un objet disparaît et réapparaît dans un scène naturelle complexe, les auteurs ont mis en évidence que l'on est

beaucoup plus rapide pour détecter ce changement s'il affecte un animal qu'un autre type d'objets. Même les voitures qui constituent de nos jours une source de danger importante et qui par leur vitesse de déplacement doivent être détectées rapidement, ne bénéficient pas de cet avantage. Les auteurs concluent donc sur notre capacité « innée » à détecter les stimuli biologiquement importants. L'homme possèderait donc une prédisposition à détecter les animaux dans une image. Pourquoi pas les singes ? Les singes pourraient ainsi -sans en avoir côtoyé directement- posséder une connaissance innée de ce qu'est un animal. Et dans ce cas, la question de traitement dans un mode de confusion ou d'équivalence entre objet réel et photographie se pose.

Une catégorie d'animaux dont la détection rapide paraît particulièrement importante est celle des prédateurs. Détecter rapidement un prédateur permet d'avoir immédiatement la réaction appropriée permettant la survie de l'animal. C'est un critère adaptatif qui mérite, si possible, d'être transmis aux générations futures. Les prédateurs ancestraux par excellence des primates sont les serpents (Isbell 2006). Effectivement, on constate que les hommes adultes (Ohman et al. 2001), les enfants (LoBue et DeLoache 2008) ainsi que les macaques japonais (Shibasaki & Kawai 2009) sont plus rapides pour détecter des images de serpents que des images de stimuli neutres comme des champignons ou des fleurs. Il semble que ce biais particulier vers les serpents soit au moins partiellement inné. En témoigne la prédisposition des macaques élevés en laboratoire (Cook et Mineka 1989) et des jeunes enfants (dès 7 mois, DeLoache et LoBue 2009) à acquérir cette peur des serpents en visionnant des vidéos où les serpents sont associés à des réactions de peur de leur congénères.

Dans une expérience introduite en Annexe 1, nous nous sommes intéressés à la façon dont nos singes traitent les serpents. Nos deux singes ont été entraînés avec des images extrêmement variées d'animaux représentant toutes les grandes classes (reptiles, mammifères, oiseaux, poissons, insectes). Comme présenté dans l'information supplémentaire de l'article 1, ils sont capables de généraliser à de nouvelles images de chacun de ces grands groupes et les considèrent donc tous comme appartenant à la méta-catégorie animale. Néanmoins, nous n'avons jamais analysé le rôle d'une espèce ou d'un genre particulier d'animal au sein de ces

grandes classes. Un article récent de notre groupe (Girard et Koenig-Robert 2011) suggère que dans une tâche de saccade vers la catégorie animale les macaques ont tendance à avoir de meilleures performances sur les reptiles. De même les humains, dans une tâche de catégorisation manuelle animal/non animal, identique à celle passée par nos singes, sont plus précis et plus rapides sur les reptiles que sur les autres groupes d'animaux (Delorme et al. 2011). Qu'en est-il de nos deux macaques? Quelles sont leurs réactions lorsqu'on leur présente de nouvelles images de serpents?

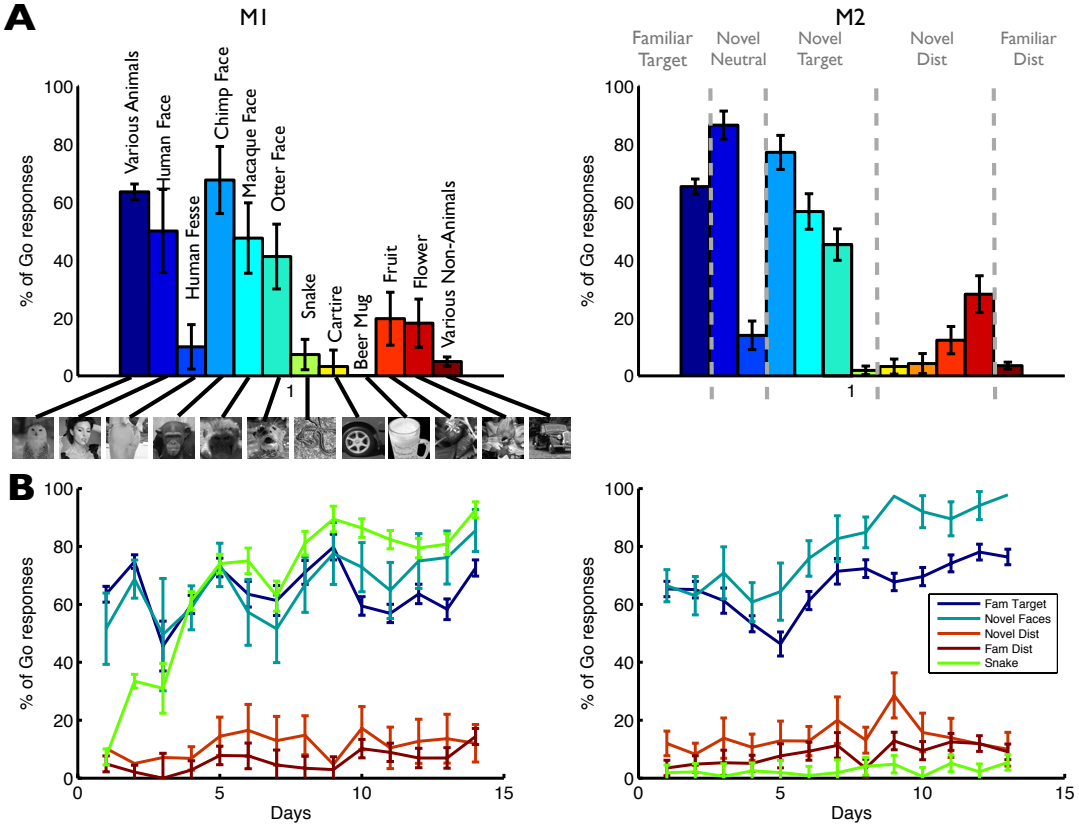

Figure 5: Les images de serpents induisent des réponses comportementales particulières Pourcentage de réponses animales go (+/- intervalle de confiance à 95 %) pour les deux singes (M1 à gauche et M2 à droite) durant la première séance d'introduction des nouveaux stimuli (A) et durant l'ensemble de l'expérience (B). En A, chaque barre de couleur correspond à une catégorie de stimuli, dont la nature est illustrée par un exemple et décrite par un nom, à gauche. Leur statut de cible (récompensé pour une réponse go), distracteur (récompensé pour une réponse no go) ou neutre (les stimuli humains étaient toujours récompensés) ainsi que leur caractère familier ou nouveau est indiqué à droite. En B, chaque point correspond à un jour. En bleu foncé sont représentées les réponses go sur les images d'animaux cibles familières, en rouge foncé sur les images non animales distractives familières, en turquoise les réponses sur les nouvelles images animales cibles (chimpanzé, macaque, loutre), en orange les réponses sur les nouvelles images non animales distracteurs (roue de voiture, bière, fruit et fleur) et en vert les réponses sur les serpents.

La Figure 5A présente le pourcentage de réponse go sur des nouveaux stimuli bien spécifiques que sont les visages humains, macaques, chimpanzés et les serpents. Les deux singes répondent significativement moins sur les serpents que sur les autres types d'animaux. Ils répondent même moins que sur certains

distracteurs comme les fleurs ou les fruits. La Figure 5B montre comment ces performances évoluent avec l'apprentissage et le protocole de renforcement, qui récompense chaque réponse go sur les serpents par une goutte d'eau, et « punit » chaque réponse no go sur ces mêmes stimuli en réaffichant l'image pendant 3 secondes. On observe que le singe 1 va répondre sur les serpents aussi bien que sur les autres animaux au bout de 3 jours. Le singe 2, pendant les 13 jours d'expérience, ne produira jamais plus de 5% de réponses go sur les serpents. Une performance bien inférieure à celle obtenue sur les autres cibles animales.

Il y a donc un refus de répondre pour les deux singes sur les images de serpents lors de la première session de test qui va perdurer chez un des deux singes tout au long des 13 jours d'expériences. Il faut savoir que dans ce protocole on demande aux singes de toucher l'image cible sur l'écran tactile. Les singes semblent démontrer une réaction d'inhibition ou d'évitement innée, comme s'ils étaient face à un vrai serpent. C'est la définition même du stade de confusion. Un des deux singes reste clairement dans ce mode de confusion durant toute l'expérience. Le singe 1 dépasse le stade de confusion et apprend à associer les serpents à la catégorie animale, sur laquelle il faut répondre pour être récompensé, ce qui laisse penser qu'il pourrait atteindre un niveau d'équivalence avec l'apprentissage (Fagot et al. 2010).

2.3.4 Concept naturels : plus loin dans l'abstraction ?

Les concepts naturels qui semblent demander le plus fort niveau d'abstraction, parce qu'ils ne partagent, par définition, pas de caractéristiques physiques, sont certainement les concepts multimodaux. Les singes sont tout à fait capables d'accéder à ce genre de concept notamment au niveau de l'espèce comme à celui de l'individu. En utilisant, un paradigme de priming multimodal, Martin-Malivel et Fagot (2001) montrent que les babouins engagés dans une tâche de discrimination de cris de conspécifiques et de voix humaines, répondent plus rapidement si le son est précédé d'une image congruente au niveau de l'espèce. Cet effet de priming ou de facilitation visuels dans une catégorisation auditive témoigne de l'existence d'une représentation amodale du concept de babouin ou du concept d'être humain.

Une étude plus récente, sur les macaques rhésus, suggère leur capacité à former spontanément un concept d'individus conspécifiques ou humains (Silwa et al. 2012). En enregistrant le temps de fixation oculaire, ils rapportent que les singes regardent significativement plus longtemps la photo du visage d'un individu s'ils viennent juste d'entendre sa voix.

Enfin, un concept naturel spécial est celui du concept de soi témoignant d'une conscience de soi chez l'animal. Nous avons déjà parlé de la capacité des singes à se reconnaître dans un miroir (Gallup 1970). Ce comportement a été retrouvé chez toutes les espèces de grands singes (Orang-Outan, Bonobo, Gorille), en revanche l'existence de cette capacité chez les autres primates reste sujette à controverse. Ainsi si des études récentes montrent l'existence de prémices de comportement orienté chez les tamarins, les capucins moines et différentes espèces de macaques, ces comportements ne sont jamais aussi développés que ceux rapportés chez les grands singes et sont souvent affectés par des biais expérimentaux (Anderson et Gallup 2011). Avec le concept de soi, on toucherait donc du doigt une fonction cognitive qui serait non pas le propre de l'homme mais d'un petit nombre d'espèces animales dont les grands singes (homme compris), certains cétacés, certains corvidés et peut-être les éléphants.

2.3.5 Concepts abstraits basés sur des règles

La capacité à former des concepts relationnels chez les primates non-humains a longtemps été considérée comme un critère distinctif entre singes et grands singes (humains inclus, Premack, 1976; Thompson et Oden, 2000). Effectivement, si les babouins sont capables de réaliser une tâche identique/différent, c'est uniquement avec un grand nombre d'icônes (Fagot et al. 2001). Ils ne réussissent la tâche en n'utilisant que deux icones seulement, uniquement si elles sont présentées côte à côte et séparées par un très faible intervalle (Fagot et Parron 2008), laissant penser qu'alors les singes n'utilisent pas une règle d'identité ou de différence, mais traite la paire d'icônes comme un seul stimulus.

Une étude récente vient éclairer ce débat en montrant que sur une population de 26 babouins, 6 sont capables de réussir une tâche d'appariement différé relationnel (Fagot et Thompson 2011). Certains singes sont donc bien capables de maîtriser ce genre de règle relationnelle analogique.

La même équipe a récemment publié des travaux encore plus surprenants montrant que les babouins étaient capables de distinguer des mots de non mots (Grainger et al. 2012). Les auteurs défendent ainsi l'existence de compétences orthographiques chez un animal.

Enfin, un concept abstrait auquel nous avons souvent recours est celui de quantité. Les singes sont aussi capables d'accéder à la notion de quantité et même de numérosité. Effectivement, ils peuvent ordonner correctement des stimuli contenant de 1 à 9 items (Brannon et Terrace 1998).

3 Neurosciences comparées de l'abstraction visuelle

3.1 Entrée en matière

Dans le chapitre précédent nous avons présenté un éventail d'arguments comportementaux qui poussent à croire que de nombreux animaux peuvent former des concepts : représentations mentales d'objets, d'individus ou de catégories. Des espèces très éloignées d'un point de vue phylogénétique sont capables de reconnaître leurs congénères, leurs parents, de former diverses catégories naturelles à un niveau basique ou superordonné ou encore des catégories plus abstraites basées sur des règles relationnelles ou analogiques.

Il semblerait que finalement un certain nombre de facultés conceptuelles soient essentielles pour la survie de n'importe quelle espèce sur cette terre. A cause de pressions du milieu similaires, des espèces, même très éloignées, possédant des systèmes nerveux très différents, auraient convergé vers un même jeu de mécanismes de conceptualisation.

Cette idée d'adaptation convergente est généralement opposée à celle d'homologie, selon laquelle on retrouve, chez des espèces proches par la phylogénie, un même organe provenant d'un ancêtre commun, pouvant ou non sous-tendre la même fonction. L'homologie témoigne de l'origine commune des espèces et a permis à Darwin de postuler la théorie de l'évolution par la sélection naturelle. Ce principe fondamental de l'organisation du vivant a été tellement bien confirmé par la découverte de l'ADN et la génétique, qu'on l'applique sans complexe à tous phénomènes biologiques, y compris aux comportements et à la cognition.

Néanmoins, malgré les efforts de la cognition comparée, les preuves de l'existence de réelles homologies cognitives restent faibles, principalement pour deux raisons. D'abord parce que la cognition comparée ou cognition animale s'attache plus à montrer l'existence de telle ou telle fonction cognitive chez une espèce, plutôt qu'à la comparaison rigoureuse et méticuleuse de cette fonction à travers les espèces, seul moyen pour inférer l'existence de mécanismes cognitifs d'origine commune.

L'autre facteur qui rend la découverte d'homologies cognitives difficile est le substrat qui supporte la cognition : le système nerveux central, dont la structure et l'organisation sont extrêmement complexes. Prouver l'origine commune d'un traitement neuronal entre deux espèces est infiniment plus compliqué que montrer celle du cubitus par exemple. Or une réelle compréhension de l'évolution de la cognition, comme Darwin a pu comprendre l'évolution du vivant en général, ne pourra se faire qu'en étudiant aussi l'organe qui supporte la fonction : le cerveau. Si l'on veut comprendre l'origine et l'histoire de la cognition, nous avons besoin de neurosciences comparées.

Au niveau le plus basique et certainement le plus populaire, on peut comparer le cerveau en entier, sa taille, sa forme, sa « gyrification » (Figure 6). Une des premières approches des neurosciences comparées a été d'essayer d'expliquer une complexification des comportements et fonctions mentales par une augmentation du volume du cerveau, ou mieux par le coefficient d'encéphalisation d'une espèce (taille du cerveau rapportée à la taille du corps entier de l'animal). Ce coefficient d'encéphalisation, en plaçant l'homme tout au sommet de la pyramide, suivi de près par le dauphin, le chimpanzé, puis le macaque et les autres mammifères a satisfait les biologistes du XIXème siècle (Dubois 1897) et est encore largement utilisé dans les traités de vulgarisation actuels.

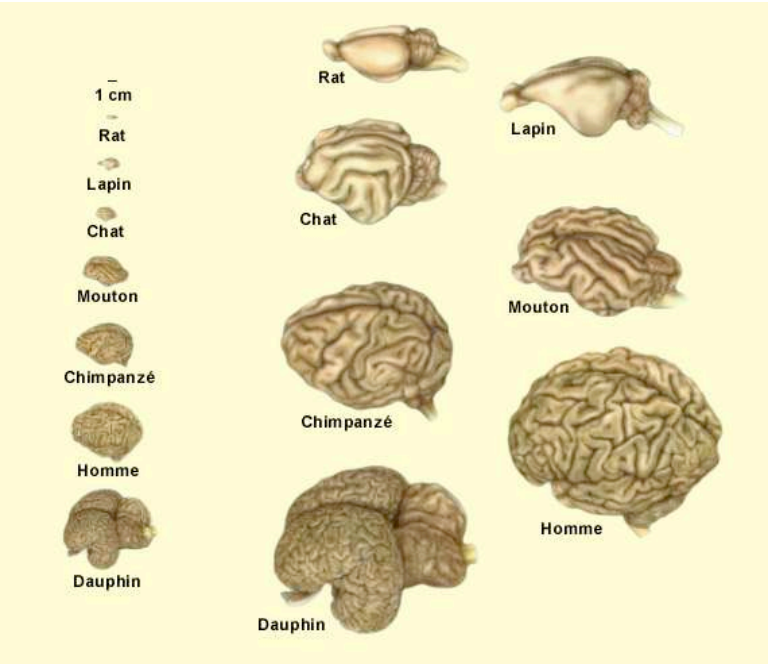

Figure 6 : Exemples de cerveaux de mammifères

Néanmoins, à l'heure du séquençage de l'ADN, de la compréhension du fonctionnement synaptique, de l'enregistrement du neurone unitaire ou de l'IRM haute résolution, les neurosciences comparées se doivent de développer une approche beaucoup plus raffinée et d'intégrer les différents niveaux de fonctionnement du cerveau qu'offrent ces nombreux outils. Le sujet est vaste et il reste beaucoup à faire. Dans le cadre de cette thèse nous nous contenterons de chercher et de comparer le système neuronal impliqué dans l'élaboration de concepts visuels chez l'homme et le macaque. Nous adopterons néanmoins une méthode originale puisque, même si certains ont commencé à s'intéresser à ce sujet en utilisant l'IRM fonctionnelle (Orban 2004, Tsao et al. 2008) ou en combinant IRM fonctionnelle chez l'homme et enregistrements unitaires chez le singe (Kriegeskorte et al. 2008), c'est la première fois que l'on essaie de comparer des enregistrements neurophysiologiques de même nature, chez des sujets en comportement.

Les neurosciences comparées de l'abstraction visuelle dans lesquelles nous avons voulu nous plonger ici ne sont pas anodines non plus puisque, comme nous l'avons expliqué au début de ce mémoire, étudier la formation des représentations ou concepts visuels est un moyen opérationnel de s'attaquer à une étude rigoureuse de la pensée, en en décortiquant les briques élémentaires. Avant les substrats neuronaux des concepts visuels, commençons par décrire et expliquer la remarquable circuiterie neuronale qui nous permet d'extraire du sens de la lumière : le système visuel.

3.2 Le système visuel

La plupart des résultats décrits ci-dessous ont été obtenus chez des animaux (chats et macaques rhésus) que l'on prend ironiquement comme de bons modèles du système visuel humain, lorsqu'il s'agit de pratiquer des études électrophysiologiques lourdes et invasives, mais à qui on refuse bien souvent la possession d'émotions ou de fonctions cognitives similaires aux nôtres. Le système visuel que nous allons décrire ici n'est donc pas celui de l'homme, sur lequel on n'a principalement que des connaissances globales, mais celui des mammifères ou primates en général. Néanmoins, de par les ressemblances anatomiques et cytoarchitectoniques qui ont pu être observées post mortem chez l'homme, on admettra au moins que jusqu'au cortex visuel primaire, le système visuel humain doit aussi fonctionner comme décrit ci-dessous.

Au commencement il y a la lumière, une onde électromagnétique composée de particules, les photons, et dont le spectre visible oscille avec une longueur d'onde comprise environ entre 400 et 800 nm. Les pigments des objets absorbent sélectivement certaines longueurs d'ondes émises par la lumière tandis qu'ils réfléchissent et renvoient la lumière correspondant aux autres longueurs d'ondes non absorbées. Ce sont ces ondes réfléchies qui sont captées par la rétine et que nous percevons. Percevoir un stimulus visuel débute donc par l'arrivée des photons réfléchis par le stimulus, sur la rétine. Après avoir traversé la cornée, l'humeur aqueuse, le cristallin et l'humeur vitrée, ils viennent activer les photorécepteurs de la rétine assurant ainsi la transduction de l'information lumineuse en messages nerveux.

La rétine est composée de différents types de cellules nerveuses: les photorécepteurs, appelés cônes ou bâtonnets, des cellules relais, appelées cellules bipolaires, des cellules de projections, appelées cellules ganglionnaires, (elles-même divisées en deux classes: P pour parvocellulaire et M pour magnocellulaire) et des interneurones, appelés cellules amacrines et horizontales.

Les photorécepteurs ont chacun une sensibilité propre à la lumière. Les bâtonnets, les plus nombreux (environ 120 millions), sont très sensibles. Leur activité est souvent saturée en condition d'éclairement normal. Du fait de cette importante sensibilité, ils participent plutôt à la vision crépusculaire et nocturne. Au contraire, les cônes, moins nombreux (5 millions de cellules environ) ont une sensibilité plus faible et participent donc plutôt à la vision diurne, lorsque la luminosité est normale. Les cônes ont une sensibilité spectrale différente des bâtonnets, leur permettant de participer à la vision des couleurs tandis que les bâtonnets ne permettent qu'une vision achromatique, en noir et blanc. La répartition des photorécepteurs n'est pas uniforme. La concentration des cônes est maximale au centre de la rétine, appelée fovéa (centre de la macula), permettant ainsi une acuité visuelle importante au centre du champ visuel. Plus on s'éloigne de la fovéa, plus le nombre de cônes diminue et plus l'acuité visuelle sera faible. Chaque photorécepteur code un point précis du champ visuel et l'information visuelle est transmise aux cellules ganglionnaires. Les cellules ganglionnaires, quant à elles, recevant des signaux de plusieurs cellules nerveuses, vont en quelque sorte « intégrer » l'information provenant des photorécepteurs.

A ce stade, il est important d'introduire la notion de champ récepteur d'un neurone qui constitue la région du champ visuel qui, lorsqu'elle est stimulée, va moduler la réponse du neurone. Ainsi le photorécepteur a un champ récepteur plus petit, tandis que les cellules ganglionnaires, qui vont être activées par un grand nombre de photorécepteurs (sauf à la fovéa qui respecte une organisation linéaire), auront un champ récepteur un peu plus grand et ainsi de suite. Plus l'information progresse dans la hiérarchie visuelle, plus les champs récepteurs des cellules vont être de grande taille.

Les axones des cellules ganglionnaires constituent le nerf optique, les potentiels d'action générés vont se propager vers le corps genouillé latéral (CGL, noyau du thalamus). Au sein du CGL, deux types de couches cellulaires ont été identifiées : quatre couches parvocellulaires qui reçoivent des inputs des cellules ganglionnaires de type P et deux couches magnocellulaires recevant des informations transmises par les cellules M (il existe aussi des cellules koniocellulaires dont le rôle est moins clair). Les cellules P, ont de petits champs récepteurs et répondent aux hautes fréquences spatiales, la voie parvocellulaire est généralement associée à un traitement fin mais lent de l'information visuelle. Au contraire les cellules M, avec de grands champs récepteurs et des réponses aux basses fréquences spatiales, projettent sur un système permettant un traitement rapide mais grossier et achromatique de l'information.

Les informations traitées au niveau du CGL vont ensuite être transmises vers le cortex visuel primaire (V1). Le cortex V1 (aire 17 selon l'architecture cérébrale décrite par Brodmann) est situé sur la face interne du lobe occipital de chaque hémisphère, enfoui à l'intérieur de la scissure calcarine chez l'homme. V1 est organisé de manière rétinotopique, c'est-à-dire que cette région reconstruit une « carte topographique », point par point, du champ visuel à partir des informations en provenance des cellules ganglionnaires. Les études électrophysiologiques pionnières d'Hubel et Wiesel, chez le chat anesthésié, ont montré que les neurones de V1 répondent à des barres lumineuses orientées, en mouvement ou non. Depuis le cortex visuel primaire, le signal est ensuite envoyé à différentes aires visuelles associatives interconnectées (travaux pionniers chez le singe: Van Essen et al. 1992) qui vont traiter les informations pour en extraire à chaque étape les différents indices de la scène visuelle permettant l'analyse de chacune de ses caractéristiques.

On dissocie deux voies principales du traitement de l'information visuelle mises en évidence dans le cerveau du singe par Ungerleider et Mishkin en 1982 (voir aussi Mishkin, Ungerleider, Macko, 1983 et

Figure 7).La voie visuelle dorsale (vers le lobe pariétal) ou voie du « où » qui code les informations relatives au mouvement, à la position dans l'espace permettant ainsi la localisation des objets dans notre environnement. Cette voie inclut l'aire MT (pour aire médiale temporale). La voie visuelle ventrale (vers le lobe temporal) ou voie du « quoi » qui depuis le lobe occipital, se prolonge à la surface ventrale du cerveau, incluant V4, jusqu'au lobe temporal et permet la perception des formes et des couleurs, la reconnaissance et l'identification des objets (Dicarlo et al. 2012). Cette dissociation voie ventrale/voie dorsale a également été décrite chez de nombreux animaux et chez l'homme, grâce aux études lésionnelles ou à l'imagerie cérébrale chez les sujets sains (Haxby et al., 1991).

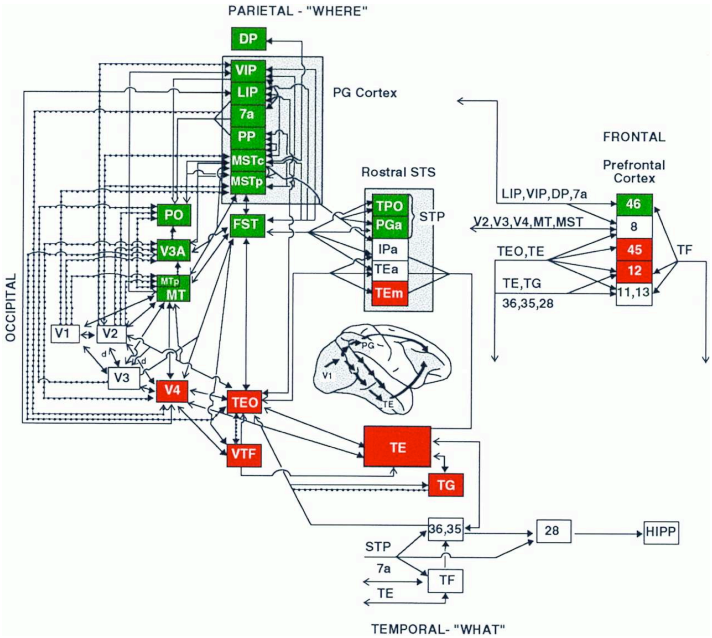

Figure 7 : Les aires visuelles chez le singe (d'après Ungerleider et al. 1998). En rouge les aires de la voie ventrale (voie du « quoi ? ») en vert les aires de la voie dorsale (voie du « où ? »). Les flèches pleines indiquent des connexions ascendantes (« feedforward »), les flèches vides (<-) des connexions descendantes (« feedback »).

Dans ce mémoire de thèse, nous allons nous focaliser sur la voie visuelle ventrale du fait de son implication dans le traitement et la représentation des objets visuels.

3.3 Matériel et méthodes : Comment étudier les bases biologiques des représentations mentales ?

Quel système d'enregistrement de l'activité cérébrale est le plus à même de capturer l'émergence d'une représentation mentale à partir de son substrat neuronal ?

Les sciences de l'imagerie biologique ont fait de remarquables progrès ces trente dernières années, il existe maintenant toutes sortes de gros équipements et de petits senseurs pour capturer la magie du vivant en fonctionnement. Pour essayer de répondre à une question spécifique en biologie, il convient donc de faire le meilleur choix parmi toutes ces possibilités. Ce choix s'effectue selon trois critères :

- Quelle est la nature physique des signaux que l'on souhaite étudier (pression, température, voltage électrique …)?
- Quelle est la dimension temporelle du phénomène que l'on souhaite étudier (ms, heure, jours …)?
- Quelle est l'échelle spatiale, ou niveau d'intégration, qui nous intéresse (la molécule, la cellule, l'organe, l'individu, le groupe d'individus…) ?

Après avoir brièvement décrit les principales méthodes d'imagerie cérébrale, nous verrons qu'il est pour nous pertinent de se focaliser sur les techniques dites d'électrophysiologie, puisqu'elles permettent à la fois d'enregistrer l'activité électrique résultant directement du fonctionnement neuronal et de fournir une résolution temporelle suffisante pour étudier la dynamique des représentations.

Il s'agira ensuite de décider de la résolution spatiale ou niveau d'intégration neurale pertinent. Nous nous arrêterons sur les techniques d'enregistrements intracrâniens car elles présentent de nombreux avantages : (1) elles échantillonnent le cortex de façon objective, (2) elles témoignent de l'activité d'un large réseau neuronal en fournissant un enregistrement simultané de plusieurs territoires corticaux, (3) elles peuvent être utilisées à la fois chez l'homme, dans le cas d'épilepsies pharmaco-résistantes, et chez le singe; elles fournissent par conséquent un outil idéal pour la comparaison des deux espèces.

Il ne suffit pas d'enregistrer le bon signal, il faut aussi être capable d'utiliser le bon « décodeur » pour pouvoir le traduire et l'interpréter. Nous verrons que l'apprentissage artificiel est un outil idéal pour lire les représentations, puisqu'il

permet d'extraire le maximum d'informations d'un signal multivarié, comme celui fourni par les enregistrements intracrâniens, et ceci essai par essai.

Enfin, nous évaluerons dans cette partie "méthodes", la validité de ces systèmes d'enregistrement et d'analyses en l'appliquant au décodage des représentations formées durant la catégorisation de visages dans la voie visuelle ventrale chez le singe.

3.3.1 Comment mesurer l'activité cérébrale ?

Mesures électriques directes vs mesures hémodynamiques indirectes

L'activité corticale mise en jeu lors de la réalisation d'une tâche cognitive peut être évaluée par deux types de mesures :

- Des mesures indirectes de l'activité neuronale. C'est le cas de l'imagerie par résonance magnétique fonctionnelle (IRMf) et de la tomographie à émission de positron (TEP), deux techniques qui reposent sur le fait que: le traitement de l'information par le cerveau va augmenter localement le flux sanguin, pour répondre à la demande en énergie des neurones en activité. Ce sont ces variations hémodynamiques qui sont quantifiées, et permettent d'inférer les territoires corticaux impliqués dans le traitement de l'information en question. Ces méthodes ont l'avantage d'imager la totalité du cerveau et ainsi de donner une idée globale du réseau d'aires cérébrales impliquées dans le traitement de l'information.
- Des mesures directes de l'activité électrique neuronale. Un neurone en activité reçoit de multiples courants électriques excitateurs et inhibiteurs appelés respectivement PPSE (potentiel post synaptique excitateur) et PPSI (potentiel post synaptique inhibiteur) dont l'intégration va provoquer, si l'excitation électrique est suffisante, l'émission d'un potentiel d'action (PA). Tous ces phénomènes électriques, reflétant l'activité d'un ou plusieurs neurones, sont mesurables à différentes échelles. L'ensemble des méthodes visant à enregistrer ces activités électriques est appelé électrophysiologie. Ces méthodes ont pour gros avantage de refléter réellement l'activité électrique produite par les neurones, et surtout de posséder une résolution temporelle extrêmement fine (de l'ordre de la ms), à l'échelle des actes cognitifs qui nous intéressent.

Méthodes d'électrophysiologie chez le sujet en comportement: Du potentiel d'action à l'électroencéphalogramme

De nombreuses études sur le fonctionnement du système visuel ont été effectuées, par commodité, sur des animaux anesthésiés ou simplement entraînés à passivement fixer l'écran pendant que des stimuli leur sont présentés. Or si ces études permettent d'avoir une idée de la mécanique du système visuel, elle ne permettent absolument pas de savoir si les phénomènes observés sont réellement liés à la perception du sujet. Pour étudier les traitements neuronaux impliqués dans la perception du sujet, il convient donc d'enregistrer l'activité neuronale pendant que celui-ci est en train d'effectuer une tâche contrôlée et de lier les phénomènes neuronaux aux réponses comportementales. Quelles méthodes d'électrophysiologie peuvent être utilisées chez l'animal en comportement ?

Il est possible d'enregistrer l'activité d'un seul neurone en utilisant des microélectrodes : c'est l'enregistrement unitaire. Si les enregistrements unitaires intracellulaires (l'électrode transperce la membrane cellulaire) ont fourni des informations cruciales sur la physiologie interne du neurone, grâce à des expériences sur tranches ou des enregistrements sur animal anesthésié, elles sont pratiquement impossibles à mettre en œuvre sur un animal éveillé, car trop sensibles au moindre mouvement. Nous admettrons donc que le niveau minimal d'électrophysiologie pouvant être pratiqué sur un sujet éveillé est l'enregistrement unitaire extracellulaire. Cette technique a été mise en place chez l'animal en comportement pour la première fois par Evarts en 1968 (Evarts 1968). Il s'agit de faire descendre une microélectrode de quelques microns de diamètre dans la zone cérébrale d'intérêt et d'isoler un neurone répondant à la tâche que l'on fait passer au sujet.

Les microélectrodes ont été développées pour enregistrer les potentiels d'action (PA) émis par un ou plusieurs neurones. Un PA est un phénomène discret correspondant à une dépolarisation transitoire, locale, brève et stéréotypée de la membrane plasmique des neurones. Généralement d'une amplitude de l'ordre de la centaine de millivolts à l'intérieur du neurone, il ne dépasse que rarement le mV quand on l'enregistre en extracellulaire. On l'extrait à partir de l'enregistrement brut en appliquant un filtre passe haut (autour de 600, 700 Hz) et un seuil de 3-4 déviations standards par rapport

au bruit. Cette procédure a été extrêmement et majoritairement utilisée jusque dans les années 2000.

Ce n'est que récemment que les électrophysiologistes ont commencé à s'intéresser à l'information portée par l'activité continue basse fréquence enregistrée par ces mêmes électrodes, appelée potentiels de champs locaux (LFP, local field potentials) généralement obtenus par un filtre passe bas à 200 Hz. Ces potentiels de champs locaux refléteraient l'activité de plusieurs milliers de neurones. La résolution spatiale de cette activité semble varier en fonction des aires : alors que dans V1 elle serait d'environ ~250 µm autour de la pointe de l'électrode (Katzner et al. 2009), dans IT la sélectivité des LFPs est prédite par la sommation des PA sur 3 mm minimum (Kreiman et al. 2006).

Les PA signalent l'intégration de l'information émise au niveau de l'axone par le neurone. Les LFPs quant à eux, émaneraient plutôt de l'activité dendritique (ppse et ppsi) et donc correspondraient plus à l'activité entrante dans la région enregistrée (Monosov et al. 2008).

Si au lieu d'utiliser des microélectrodes (faible diamètre :~1 µm et forte impédance : 1-10 MegOhm) on a recours à des macroélectrodes (gros diamètre : >100 µm et faible impédance : <kohm), il devient alors impossible d'enregistrer les potentiels d'action, mais on accède aux potentiels de champs intracrâniens (IFP, intracranial field potential). Même si la résolution spatiale de ce genre de signal n'a pas été très bien étudiée, elle est vraisemblablement inférieure à celle des LFPs (et donc capte l'activité d'une plus grande population de neurones). Cependant dans les enregistrements intracrâniens avec des électrodes stéréotaxiques, on constate que l'activité entre deux plots d'enregistrement peut être extrêmement différente (Barbeau et al. 2007), ce qui laisse penser que, dans certains cas, la résolution de cette méthode peut être inférieure à 2 mm (distance entre 2 plots). On différencie les IFP subduraux enregistrés sous la dure-mère (comme nous le faisons avec les macro-fils chez le singe, les électrodes stéréotaxiques utilisées en sEEG ou en EcoG subdurale) des enregistrements épiduraux (les électrodes ne traversent pas la dure-mère). La première méthode fournit un signal d'un meilleur rapport signal sur bruit, car elle n'est pas soumise à la diffusion due aux méninges.

Les IFPs d'abord utilisés pour étudier les rythmes d'éveil/sommeil et attentionnels chez le rat et le chat (Rougeul-Buser et al. 1975) ont été récemment remis au goût du jour pour l'analyse en potentiels évoqués chez le singe (Peissig et al. 2007, Woodman et al. 2007). En parallèle, l'enregistrement des IFPs chez les patients épileptiques pharmaco-résistants est aussi devenu une technique de plus en plus pratiquée dans les hôpitaux pour localiser précisément le foyer épileptique.

Enfin, on peut aussi enregistrer l'activité électrophysiologique directement à la surface du crâne, c'est l'électroencéphalographie (EEG). Avec cette technique, même si on garde une très bonne résolution temporelle, la double barrière méninges et crâne provoque une diffusion importante du signal dont la source originale est difficile à localiser.

Traquer les représentations : quelle méthode d'enregistrement choisir ?

Nous éliminerons directement les méthodes reposant sur l'activité hémodynamique.
En effet, même si des tentatives ont été menées pour lier activité neuronale et signal BOLD enregistré en IRM fonctionnelle (Logothetis et al. 2001), ces interprétations sont encore sujettes à polémique. De plus, nous avons vu dans le chapitre précédent que les primates étaient capables de former un concept visuel en moins de 200 ms. Une technique pertinente pour traquer les représentations mises en jeu dans de telles tâches doit avoir une résolution de l'ordre de la milliseconde, alors que celles de l'IRMf et de la PET sont de plusieurs secondes.

Nous utiliserons donc les techniques d'électrophysiologie, mais quel type de signal enregistrer et quel niveau d'intégration serait le plus pertinent? Les microélectrodes permettent d'enregistrer l'activité d'un seul neurone, ce qui nous renseigne peu sur le fonctionnement global du cerveau. De plus, cette méthode d'enregistrement souffre d'un biais d'échantillonnage. Effectivement, on cherche, parfois pendant des heures, le « bon » neurone - celui qui répond au test que l'on a mis en place - ignorant des centaines d'autres cellules.

Avec l'augmentation de la puissance de calcul des ordinateurs et la miniaturisation des systèmes d'enregistrements il est maintenant possible d'enregistrer, non pas avec une seule électrode, mais avec une matrice d'électrodes en conservant la même

résolution temporelle. Ce type de réseau d'électrodes peut être implanté de façon chronique chez l'animal, ce qui évite les multiples violations de la barrière hémato-encéphalique nécessaires à la descente quotidienne de microélectrodes et donc le risque d'infections. Ces enregistrements électrophysiologiques dit multivariés permettent d'obtenir un échantillonnage impartial ; même si les électrodes sont placées dans des aires d'intérêt, elles ne sont pas déplacées jusqu'à trouver un neurone dont les modalités de réponse sont celles recherchées par l'expérimentateur.
Des matrices possédant jusqu'à 100 microélectrodes et donc permettant potentiellement d'enregistrer une centaine de neurones en même temps, on été créées pour étudier le cortex moteur. Au début de mon travail de thèse nous avons essayé de transférer cette technique au système visuel. Néanmoins cet essai de 6 mois a été très peu concluant notamment sur la capacité à obtenir un signal informatif et stable au fil des jours d'enregistrement. Heureusement, aujourd'hui d'autres ont réussi ce transfert, notamment une prestigieuse équipe d'Harvard (Ray et Maunsell 2010).

Ne disposant pas de tels outils fonctionnels, nous avons eu recours à une technique alternative, peu utilisée actuellement dans les neurosciences cognitives chez le singe, consistant à implanter plusieurs macrofils chroniquement dans les aires corticales d'intérêt. Cette méthode possède plusieurs avantages :

- elle jouit d'une résolution temporelle de l'ordre de la ms suffisamment précise pour analyser finement les processus neuronaux mis en jeu dans nos tâches de catégorisation rapide
- elle nous permet d'enregistrer simultanément l'activité de plusieurs territoires corticaux, répartis dans la voie visuelle ventrale, et fournit par conséquent un outil adéquat pour traquer l'émergence de représentations corticales visuelles même distribuées.
- elle est similaire aux techniques d'enregistrement intracrâniens utilisées chez le patient épileptique et donc va nous permettre pour la première fois de tenter la mise en évidence d'homologies électrophysiologiques entre le singe et l'homme.

3.3.2 Nouvelles méthodes d'analyses pour décoder les représentations mentales

L'électrophysiologie nous donne la matière. Mais une fois cette matière acquise il est temps de ranger la blouse et de faire marcher la machine à décrypter les pensées. Il faut montrer qu'il existe bien, dans ces fluctuations de potentiels électriques émanant de l'usine neuronale, une représentation intérieure du monde extérieur. Encore une fois quelle démarche adopter? Idéalement on aimerait une méthode qui puisse :

- tirer un maximum d'information des enregistrements neuronaux, en prenant en compte l'enregistrement multi-sites par plusieurs électrodes
- quantifier cette information avec précision pour chaque présentation de stimuli
- évaluer la dynamique de traitement de cette information
- permettre de faire un lien entre traitement neuronal et réaction comportementale à chaque essai.

Comme nous allons le voir, si les techniques classiques d'analyse du signal ne remplissent pas tous ces critères, le champ émergent consistant à transférer les techniques d'apprentissage artificiel et de classification aux neurosciences semblent être l'outil idéal pour décoder les représentations neuronales. C'est-à-dire pour tenter de lire dans les pensées.

Les PEVs

L'analyse classique des potentiels de champs dans un protocole de présentation visuelle consiste à isoler des portions de signal autour de la présentation du stimulus, puis à les moyenner, à chaque pas de temps, sur un grand nombre d'essais afin d'éliminer le bruit qui fluctue d'un essai à l'autre. Il en résulte des potentiels évoqués visuels (PEV).

Cette méthode a été extrêmement utilisée en EEG ces 20 dernières années, elle a permis de mettre en évidence certaines composantes définies par leur positivité (p) ou négativité (n) et par la latence à laquelle elles atteignent leur amplitude maximale. Ainsi les PEVs en scalp présenteraient un premier pic aux alentours de 100 ms, la P100 vraisemblablement impliquée dans la détection d'objets visuels, puis un pic négatif la N170 spécifique chez l'homme du traitement des visages, suivi d'une N250

dont l'amplitude serait reliée à la familiarité. On parlera aussi de P300 traduisant la surprise du sujet face à un stimulus déviant extrêmement utilisé dans les interfaces cerveau machine non invasive et d'une N400 impliquée dans le traitement sémantique.

D'autres auteurs se limitent à donner une latence minimale de traitement nécessaire pour la fonction cognitive qui les intéresse, c'est le cas de Thorpe et al. qui rapportent une latence de traitement des catégories visuelles à 150 ms post-stimulus onset en démontrant que les PEV évoqués par deux catégories de stimulus divergent à cet latence (Thorpe et al. 1996). Si ces méthodes ont été extrêmement efficaces pour délimiter les temps de traitement corticaux associés aux différents mécanismes cognitifs, elles ne permettent pas d'accéder aux contenus mentaux, essai par essai. C'est en revanche le but de l'apprentissage automatique ou analyse multi-variée appliqué aux enregistrements multi-senseurs.

Comment lire dans les pensées : l'apprentissage automatique appliqué à l'activité neuronale multi-électrodes

Depuis une dizaine d'années la combinaison des méthodes d'imagerie cérébrale et des algorithmes d'apprentissage par ordinateur a permis de s'attaquer sérieusement à un vieux rêve de l'humanité, celui de lire dans les pensées. Il est intéressant de noter que c'est en voulant créer une pensée artificielle que l'homme s'est doté d'outils lui permettant de décrypter la pensée naturelle. En effet, c'est l'avènement des ordinateurs et de l'intelligence artificielle qui a permis de développer ces algorithmes d'apprentissage automatique. Leur rôle, à l'instar du cerveau, est d'extraire un sens à partir de données brutes pouvant posséder de nombreuses dimensions. La beauté et l'efficacité de ces algorithmes reposent sur l'extraordinaire capacité de calcul des ordinateurs. Alors qu'il est difficile pour nous de visualiser et d'interpréter des données en trois dimensions, les ordinateurs sont capables d'en traiter des milliers. Cette capacité de calcul va leur permettre de tester, par simulation, un très grand nombre de possibilités en un temps minimum et ainsi de pouvoir proposer une solution optimale à un problème donné.

On pose généralement deux types de problèmes à ces algorithmes (1) retrouver dans les données des classes connues au préalable, on parlera *d'apprentissage supervisé,*

(2) extraire une certaine organisation des données, sans classe préalablement déterminée, c'est *l'apprentissage non supervisé*.

Ici, nous aurons uniquement recours à l'approche supervisée. Les données brutes que l'on utilisera seront les potentiels de champs enregistrés sur l'ensemble, ou un sous-groupe d'électrodes, les classes que nous souhaitons « lire » dans cette activité sont celles des catégories d'objets visuels présentées au sujet.

A ce stade, il est nécessaire d'introduire quelques termes techniques que nous illustrerons pour les deux applications de l'apprentissage artificiel décrites dans cette thèse : l'analyse multi-variée en neurosciences et la vision par ordinateur. Les données peuvent être vues comme une matrice à deux dimensions. La première dimension (lignes) correspond aux *échantillons* ou *essais* ; dans le cas de notre protocole de catégorisation visuelle, il s'agira d'un essai pour le sujet, lié à la présentation d'une image. Dans le cas des modèles de vision par ordinateur, il s'agira d'une image. La deuxième dimension est celle des *mesures* ou *caractéristiques* (« features »), c'est le signal multi-électrode associé à chaque essai. En neurosciences, il s'agira de caractéristiques extraites du signal neuronal à travers les différents voxels ou électrodes. En vision par ordinateur, on parlera de caractéristiques de l'image.

Tous les algorithmes d'apprentissage supervisé fonctionnent de la même manière. Dans une première phase, dite d'*entraînement*, l'algorithme va essayer de déterminer l'espace dans les données qui permet de séparer les différentes classes de manière optimale. La seconde phase est appelée *test*, puisqu'elle va permettre d'évaluer comment de nouveaux exemplaires vont être classifiés dans cet espace et ainsi de vérifier si la machine a bien appris à catégoriser. A l'instar de ce qui a été décrit pour la catégorisation chez l'animal dans le chapitre 2, démontrer la capacité de l'algorithme à *généraliser* est absolument crucial, pour prouver qu'il existe bien un espace permettant de séparer les classes, quels que soient les exemplaires utilisés, et non pas dépendant des exemplaires utilisés pour l'entraînement.

Il existe de nombreux algorithmes d'apprentissage automatique, différentiables par la fonction qu'ils appliquent pour séparer les différentes classes, et la méthode d'optimisation qu'ils utilisent pour y parvenir. La *fonction de séparation*, aussi

appelée *noyau*, peut être extrêmement complexe et généralement, en s'appuyant sur des hypothèses fortes de la distribution des données, elle est adaptée par des spécialistes du domaine pour chaque problème à résoudre. Dans le cas des neurosciences et en particulier pour décoder des représentations mentales, on applique le principe de parcimonie, qui stipule que pour que ces représentations soient utilisables et interprétables, par exemple pour un neurone chargé de prendre une décision, elles doivent être séparables le plus simplement possible, c'est-à-dire de manière linéaire. Dans ce cas, le but de l'algorithme est donc de trouver un ou des hyperplan(s) décrit(s) par la combinaison linéaire des mesures séparant au mieux les différentes classes. Même au sein de cette sous-classe d'algorithmes supervisés linéaires, ou classifieurs linéaires, il existe de nombreuses possibilités, néanmoins les études comparant l'efficacité de ces différents algorithmes en neurosciences, tendent à montrer qu'ils fournissent des résultats sensiblement similaires (Mehring et al. 2003; Lotte et al. 2007).

Pour des raisons d'accès libre sur internet et de rapidité de traitement, nous avons choisi d'utiliser des *machines à vecteurs de support linéaire* (« SVM for support vector machine » en anglais). Sans entrer dans les détails mathématiques de cet algorithme, la notion clé sur laquelle il repose est celle de *marge maximale*. La marge est la distance entre la frontière (hyperplan) de séparation et les essais les plus proches. Ces derniers sont appelés *vecteurs supports*. Dans les SVMs, la frontière de séparation est choisie comme celle qui maximalise la marge. Le problème est de trouver cette frontière séparatrice optimale, à partir de la série d'entraînement. Ceci est fait en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus (Fan et al. 2008).

Comment passe-t-on du SVM à la lecture de pensée ?

On dira que l'on est capable de lire dans les pensées si, à un instant t, on est capable de prédire l'état cognitif d'un sujet (ce qu'il voit, ce qu'il imagine…) sans lui demander directement. Un SVM, s'il a été entraîné à classer des états cognitifs dans un espace d'activité neuronale permettant une séparation des classes, sera capable, si on lui fournit le signal neuronal à un nouvel instant t, de prédire l'état cognitif du sujet, donc de lire dans les pensées. On parle aussi de « décodage » des pensées.

Quel type de pensées a-t-on réussi à lire chez l'homme?

Une étude rentrera dans le cadre de la lecture de pensée si elle montre la capacité d'un système à prédire un état mental en utilisant l'activité cérébrale. Théoriquement on pourrait lire dans les pensées sans utiliser les méthodes d'analyses multivariées, par exemple, si on enregistre l'activité d'un seul neurone capable de représenter différents états mentaux de manière suffisamment discriminative. En pratique, la combinaison de plusieurs neurones augmente très largement les performances de prédiction, par rapport à un décodage univarié (Quian Quirroga et al. 2007, Hung et al. 2005) et comme nous le verrons dans le chapitre suivant c'est aussi le cas pour les potentiels de champs intracrâniens.

C'est sans aucun doute la neuroimagerie chez l'homme et en particulier les études en IRMf qui ont -pour l'instant- su tirer le plus grand profit de ces nouveaux outils computationnels. La première étude marquante ayant recours à l'analyse multivariée est celle de Haxby et al. (2001), visant à étudier le codage des visages et des objets dans la voie visuelle ventrale en IRMf. Dans ce papier les auteurs sont capables de prédire la catégorie d'objets visuels présentés au sujet, en utilisant l'ensemble des voxels enregistrés dans la voie visuelle ventrale.

Cette étude pionnière fait partie de ce que l'on appellera *la lecture de pensées induites par l'environnement*, c'est-à-dire décoder une représentation qui émerge suite à une perception sensorielle directe. De la même manière, outre de nombreux travaux sur des représentations d'objets ou de catégories d'objets isolés (pour une revue voir Tong et Pratte 2012), de scènes naturelles (Peelen et al. 2009) et même d'identification de scène naturelles à travers une très grande base de données (Kay et al. 2008), on s'est aussi intéressé à décoder des traits visuels plus bas niveau comme les barres orientées, retrouvant ainsi chez l'homme, l'organisation fonctionnelle du cortex visuel primaire connu chez le singe et le chat par l'électrophysiologie (Kamitani et Tong 2005). En utilisant le même type de stimulus dans une tâche de rivalité binoculaire, Haynes et Rees (2005) confirment que, plus que les traits visuels, c'est la perception qu'en ont les sujets qui peut être lue dans le signal BOLD des aires rétinotopiques.

Les études de lecture de pensée ne s'arrêtent pas aux représentations visuelles, mais se sont aussi attaquées à la lecture des représentations olfactives (Howard et al. 2009), auditives (Ethofer et al. 2009) et même multimodales (Peelen et al. 2010).

Les différents exemples cités précédemment appartiennent à l'approche la plus commune en lecture de pensée : la classification. On prédit quelle est la classe la plus probable de la représentation observée, parmi un nombre limité de *classes* connues qui ont servi à entraîner le classifieur au préalable. L'approche dite de reconstruction, transcende la classification puisqu'elle vise à reconstruire une représentation induite par l'environnement. Au lieu de retrouver la classe d'image vue par le sujet, on va essayer de recréer cette image de toute pièce à partir de l'activité neuronale. En vision, des études d'IRMf ont montré que l'on pouvait reconstruire à travers l'activité du cortex visuel primaire (V1) des formes simples très contrastées (Miyawaki et al. 2008) mais aussi des images naturelles (Naselaris et al. 2009). En utilisant l'activité d'une aire impliquée dans le mouvement (MT) on a même réussi à reconstruire de petits films de scènes naturelles (Nishimoto et al. 2011). Sur le même principe, une étude récente vient de montrer qu'il était possible de reconstruire la parole, en utilisant l'activité enregistrée en électrocorticographie chez des patients épileptiques (Paseley et al. 2012).

En contraste avec ces études visant à décoder ou reconstruire l'expérience sensorielle, certains chercheurs se sont attaqués à lire des pensées beaucoup plus intimes. Ainsi on est capable de décoder certaines représentations formées en imagerie mentale, c'est-à-dire de prédire la catégorie d'objets que le sujet visualise intérieurement (Reddy et al. 2005). En entraînant un classifieur à dissocier les activités BOLD induites par la perception de certaines catégories, on peut significativement prédire des catégories seulement imaginées, renforçant ainsi l'idée d'un large partage des réseaux et mécanismes impliqués dans la perception visuelle et l'imagerie mentale visuelle. Des pensées particulièrement intimes, sont celles ancrées dans nos souvenirs. Il semble que les tréfonds de notre mémoire soient aussi accessibles à la lecture de pensée. Harrison et Tong (2009) ont ainsi pu accéder aux informations contenues en mémoire de travail dans les aires visuelles primaires. Polyn et al. (2005) ont montré qu'ils étaient capables de prédire le rappel ou souvenir de certaines catégories de stimuli.

Chadwick et al. (2010) ont réussi à faire la même chose mais avec des souvenirs épisodiques, personnellement vécus par le sujet.

Enfin, d'autres études se sont attaquées à décoder des activités mentales encore plus globales telles que l'attention (Kamitani et Tong 2011), la décision ou libre arbitre des sujets (Soon et al. 2008, Hampton et O'Doherty 2007), leurs intentions (Haynes et al. 2007) et même la conscience visuelle (Haynes 2009).

Cette collection d'études nous montre donc qu'il est possible d'accéder au monde mental de l'homme, qu'il émane de l'information sensorielle ou d'une génération interne. Pourrait-on appliquer ce genre d'outils pour accéder au monde mental des animaux

3.3.3 Article 2 : L'analyse multi-variée appliquée aux enregistrements intracrâniens chez le singe

Dans un premier temps, nous nous sommes appliqué à développer une méthode pour lire les représentations visuelles dynamiques chez le singe en utilisant les potentiels de champs intracrâniens enregistrés dans la voie visuelle ventrale.

Sujets, chirurgie et enregistrement

Deux singes entraînés à catégoriser des images naturelles comme contenant ou non un animal ont été équipés pour l'enregistrement de potentiels de champs intracrâniens. Il s'agit de macaques rhésus (macaca mulata) mâles, Dicky (M1) et Rouky (M2), âgés de 17 ans et 18 ans.

Les électrodes sont des fils d'aciers de 200 microns de diamètre et d'impédance à vide d'environ 7 ohm, isolées entièrement par un enrobage de Téflon, excepté au niveau de la pointe. Elles sont soudées au niveau d'un connecteur. L'ensemble (électrodes et connecteur) est fixé sur la tête de l'animal par du ciment dentaire. L'implantation est effectuée sous anesthésie générale. Pour chaque électrode, à une position préalablement calculée à partir d'images IRM 1 tesla, un petit trou d'1mm de diamètre est percé à travers le crâne. La partie terminale de l'électrode est coudée sur les derniers 2-3mm puis descendue manuellement jusqu'à rencontrer une légère résistance due à la dure-mère. Une fois la résistance dépassée, l'électrode est maintenue par le ciment dentaire. La Figure 8 A confirme l'implantation subdurale. En juxtaposant, l'IRM anatomique (faisant apparaître le cortex) pré-opératoire et le scanX post-opératoire (où figurent le crâne et les électrodes), on peut déterminer avec précision l'emplacement de l'apex de l'électrode.

L'enregistrement des potentiels de champs intracrâniens est effectué par un système Neuroscan doté d'un amplificateur Synamps, conçu pour les enregistrements de surface chez l'homme. Une électrode, frontale chez M1 et placée au niveau du sinus chez M2, sert de référence. Le signal est échantillonné à 1000Hz et filtré par passe bande de 1-200HZ.

Les singes sont assis dans une chaise pour primate et placés face à un bouton réponse infrarouge surmonté d'un écran tactile où sont affichés les stimuli. Pour récupérer les potentiels de champs, on branche simplement un câble fait maison, relié à la têtière Neuroscan, au connecteur situé sur la tête du singe. La chaise du singe et l'écran sont mis à la masse électrique pour éviter les artefacts électriques. Le premier singe M1 est tête libre pendant toute l'expérience, alors que le deuxième (M2) doit être contraint à cause des artefacts de mouvement. Il est fixé grâce à une barre en plastique maintenue sur son crâne par des vis et du ciment dentaire. Pour s'assurer de la motivation des animaux, les singes sont privés d'eau et de fruits pendant le temps de l'expérimentation. Ils n'ont accès qu'à des croquettes déshydratées. Pendant l'expérience, un système de pompe relié à un tuyau qui leur arrive à la commissure des lèvres, les récompense par une goutte d'eau avec du sirop, à chaque fois qu'ils produisent une bonne réponse.

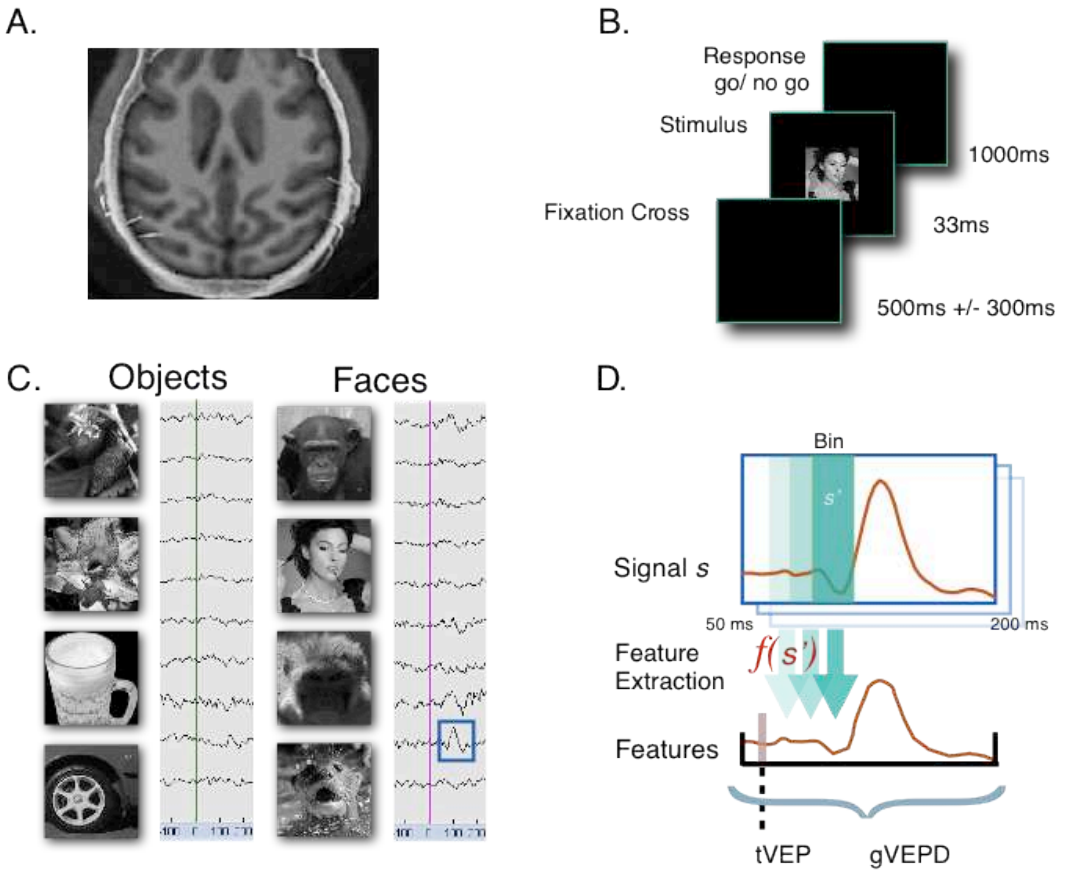

Figure 8: Méthodes expérimentales et d'analyses. Le recalage de l'IRM anatomique pré-opératoire et du scan X post-opératoire indique la position des électrodes (A). Les singes réalisent une tâche de catégorisation visages/objets dans une procédure go/no go (B et C). Le pattern multivarié d'IFPs de chaque essai est utilisé pour la procédure de décodage (C et D). Différentes caractéristiques du signal sont extraites dans les fenêtres temporelles glissantes de taille variable. Deux procédures de décodage sont utilisées : temporelle (le décodage est effectué pour chaque pas de temps indépendamment, tVEP) et globale (le décodage est effectué sur une fenêtre temporelle post-stimulus, gVEP).

Tâche

Les deux singes ont été entraînés depuis plus d'une dizaine d'années à réaliser une tâche go/no-go (Figure 8 B). Après l'affichage d'une croix de fixation centrale pendant un délai aléatoire (200-800ms), pour éviter toute prédiction ou anticipation de l'arrivée du stimulus, une image naturelle est flashée sur l'écran pendant 30 ms. Si l'image contient un animal le singe a une seconde pour lever la main et toucher l'écran tactile. La récompense est délivrée aussitôt après la réponse du singe, ce qui a pour but de le motiver à répondre le plus vite possible. Dans le cas où l'image ne contient pas d'animal, le singe doit maintenir sa main sur le bouton pendant une seconde et est récompensé à la fin de cette seconde.

En cas de mauvaise réponse, non seulement le singe ne reçoit pas de sirop mais on lui réaffiche l'image sur laquelle il s'est trompé pendant 3s, ce qui en retardant l'accès à la boisson, constitue une sorte de punition. Les animaux travaillent ad libitum jusqu'à ce qu'ils aient assez bu. Ce qui arrive généralement au bout d'1h ou 2 mais la session peut atteindre une durée de 4h. Ils réalisent en moyenne 1000 essais à l'heure.

Stimuli

Dans cette première expérience nous avons utilisé un set simple de stimuli constitués d'objets en noir et blanc en gros plan : 40 visages, cibles sur lesquelles le singe doit répondre, et 40 objets non animés, qu'il doit ignorer. Dans cette première approche, nous nous intéresserons uniquement aux représentations associées aux visages en général par rapport aux objets non animés; des distinctions plus fines (niveau subordonné) et plus larges (niveau superordonné) seront abordées dans les chapitres suivants. Cependant, on peut noter que les visages cibles sont constitués de 4 catégories : 10 visages humains, 10 visages de macaques, 10 visages de chimpanzés et 10 visages de loutres. Les distracteurs sont eux constitués de 2 catégories d'objets manufacturés (10 roues de voiture et 10 chopes de bière) et de deux catégories d'objets naturels non animés (10 fruits et 10 fleurs) (Figure 8 C). Pour éviter des différences de bas niveau entre les images, elles sont toutes égalisées en luminance moyenne et en contraste (RMS).

Analyse du signal

Les fichiers contenant les enregistrements de potentiels fournis par le Neuroscan sont importés dans Matlab grâce à une fonction de la toolbox EEGlab (Delorme et al.

2004). Ensuite, toutes les analyses sont effectuées sous Matlab par des fonctions programmées spécifiquement.

1) Prétraitement et réjection d'artefacts

Pour lire les représentations à chaque essai, le signal doit être le plus propre possible et il est nécessaire de rejeter minutieusement tous les artefacts. Dans un premier temps on applique un filtre coupe bande 45-55Hz sur le signal continu, afin d'éliminer la pollution électrique du secteur à 50Hz.

Le signal est ensuite fenêtré autour de la présentation du stimulus de -50 à 200ms. Effectivement, on a vu dans le chapitre 2 que les singes pouvaient répondre à des latences très courtes (200-220 ms après l'onset du stimulus) ; pour qu'une représentation soit utilisable avec des réponses si courtes, il faut donc qu'elle émerge dans les 100-200 premières ms. La période de -50 à 0 sert de ligne de base, pour centrer le signal autour de 0.

Pour chaque électrode et chaque essai, ce signal fenêtré est donc centré en retranchant la moyenne de la ligne de base à chaque pas de temps.

On fait ensuite une première rejection d'artefact par l'amplitude maximale absolue de voltage à 500 microvolts.

Enfin, pour compenser le fait qu'on n'ait pas de contrôle sur le regard de l'animal, on effectue une rejection d'essai par variance de la puissance visuelle. Effectivement, lorsque l'image est flashée et que le singe regarde bien l'écran on observe une brusque augmentation de la puissance du signal (somme du carré des potentiels à travers les électrodes) entre 50 et 150 ms, liée à l'arrivée massive du signal visuel dans les aires enregistrées. Etant donné que la précision des performances comportementales sur les cibles familières varient entre 90-95% par session, en appliquant le principe de précaution, on peut supputer que dans 10% des cas le singe n'as pas répondu sur des stimuli qu'il connaît parfaitement parce qu'il ne regardait pas l'écran. Pour chaque session, on enlève donc les 10% d'essais ayant la plus faible variance de puissance visuelle. On constate effectivement que sur ces essais on n'observe pas de potentiel visuel visible à l'œil nu et qu'ils ont tendance à arriver plus fréquemment en fin de session quand le singe a suffisamment bu et est moins concentré sur la tâche et l'écran.

2) Procédure de décodage

Dans l'article et les analyses supplémentaires présentés ici, on s'intéresse au décodage de représentations induites par des visages par rapport à celles induites par des objets non animés. Pour ce faire, on entraîne un SVM linéaire à dissocier la moitié des essais visages contre la moitié des essais objets en se basant sur les caractéristiques (« features ») extraites des potentiels de champs (Figure 8 D). Au cours de la phase de test, on reporte la précision de classification sur la moitié restante d'essais, dans l'espace d'entraînement. La procédure est répétée 50 fois en tirant à chaque fois de manière aléatoire les essais d'entraînement et de test. Les graphiques représentent la moyenne de la précision et l'intervalle de confiance à 95% sur ces cross-validations.

3) Extraction des caractéristiques du signal

A notre connaissance ces travaux sont les premiers tentant de lire les représentations visuelles à travers les potentiels de champs intracrâniens, chez le singe réalisant une tâche de catégorisation. Il n'y a donc pas de référence sur le type de caractéristiques du signal à utiliser. De même les questions portant sur la durée idéale de la fenêtre temporelle sur laquelle calculer ces caractéristiques et sa position temporelle par rapport au stimulus, ne sont pas définies. Dans une étude similaire chez les patients épileptiques implantés, Liu et al. (Liu et al. 2009) utilise de manière arbitraire l'étendue du signal (Max-Min) sur une fenêtre de 50 à 300 ms après présentation du stimulus.

Pour tester la meilleure combinaison possible, dans un premier temps nous avons donc évalué les performances d'un SVM linéaire pour décoder les représentations de visages et d'objets pour 9 caractéristiques du signal différentes (moyenne, intégrale, écart type, coefficient directeur, contraste de Michelson, « skewness », étendue, puissance et « kurtosis ») en utilisant 13 fenêtres temporelles d'une durée allant de 8 ms à 104 ms en augmentant la durée par pas de 8 ms, sur la première session de chaque singe (Figure 8 D et Figure 9). Afin de ne pas non plus choisir une fenêtre à une latence arbitraire après présentation du stimulus, on utilise toutes les caractéristiques extraites à partir d'une fenêtre glissante (de la taille de la fenêtre temporelle considérée) allant de 50 à 200 ms (Figure 8 C et D). On observe que la plupart des

caractéristiques et fenêtres temporelles permettent un décodage au-dessus de la chance (Figure 9). Les résultats sont assez robustes d'un singe à l'autre. Plusieurs caractéristiques et fenêtres donnent de très bonnes performances. Nous avons choisi pour l'article ci-dessous une solution simple, la moyenne sur une petite fenêtre temporelle (10 ms) ce qui permet de garder une bonne résolution temporelle. Nous verrons dans les autres chapitres que l'utilisation du potentiel brut juste recentré par soustraction de la ligne de base, fournit de très bons résultats et sera donc la caractéristique utilisée par la suite de par sa simplicité.

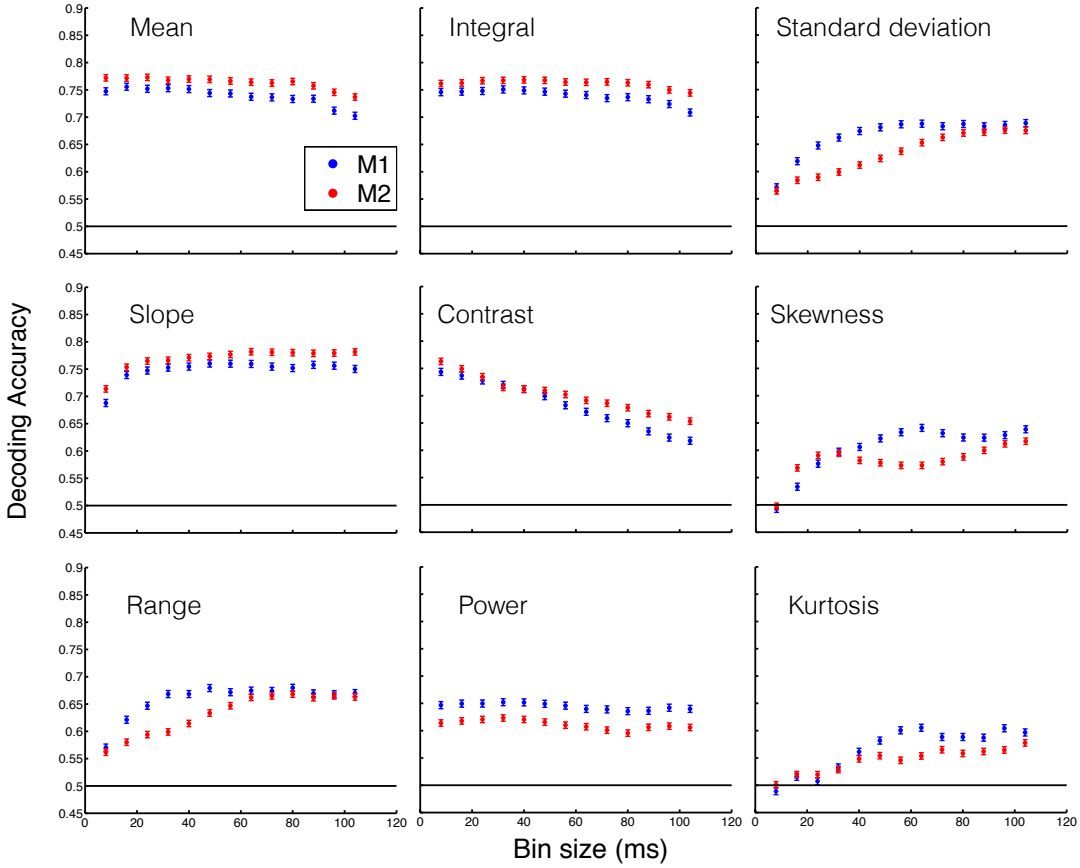

Figure 9: Précision de décodage globale visages/objets (sur la période de 0 à 200 ms) pour chaque singe (M1 en rouge et M2 en bleu) en fonction de la taille de la fenêtre d'intégration temporelle. On évalue ces performances pour 9 caractéristiques du signal : moyenne, intégrale, écart type, coefficient directeur, contraste de Michelson, « skewness », étendue, puissance et « kurtosis ».

4) Résumé de l'article

Dans ce premier article de méthodes nous avons montré que l'on peut lire chez le singe, dans les potentiels de champs intracrâniens provenant de la voie visuelle ventrale, des représentations visuelles essai par essai qui sont:

- Stables à travers les jours d'enregistrement
- Décodables très tôt après la présentation du stimulus (dès 60-80 ms après l'onset de l'image)
- Portées par un code distribué à travers les différentes électrodes et aires
- Relativement stationnaires (soutenues par le même codage pendant une centaine de ms).

On est donc capable de lire et de caractériser des représentations catégorielles, essai par essai, produites pendant la tâche de catégorisation rapide. Mais est-ce que ces représentations sont bien utilisées par l'animal pour prendre sa décision ? Ou est-ce que ce sont des épiphénomènes produits par le cerveau ? Dans le chapitre suivant, nous allons tenter de mettre en relation ces représentations et les réponses comportementales, pour évaluer leur dépendance.

ARTICLE 2: The neural dynamics of visual processing in monkey extrastriate cortex: A comparison between univariate and multivariate techniques

Maxime Cauchoix, Ali Bilgin Arlsan, Denis Fize, Thomas Serre

Sous presse dans *Lecture Notes in Computer Science*

The neural dynamics of visual processing in monkey extrastriate cortex: A comparison between univariate and multivariate techniques

Maxime Cauchoix[1,2], Ali Bilgin Arslan[3], Denis Fize[1,2], Thomas Serre[3]
[1] CNRS, CerCo, Faculté de médecine de Purpan, 31062, Toulouse, France
[2] Université de Toulouse, UPS, CerCo, 31062 Toulouse, France
[3] Department of Cognitive, Linguistic and Psychological Sciences; Brown Institute for Brain Science; Brown University, Providence RI 02912, USA

Abstract. Understanding the brain mechanisms underlying invariant visual recognition has remained a central tenet of cognitive neuroscience. Much of our current understanding of this process is based on knowledge gained from visual areas studied individually. Previous electrophysiology studies have emphasized the role of the ventral stream of the visual cortex in shape processing and, in particular, of higher level visual areas in encoding abstract category information. Surprisingly, relatively little is known about the precise dynamics of visual processing along the ventral stream of the visual cortex. Here we recorded intracranial field potentials (IFPs) from multiple intermediate areas of the ventral stream of the visual cortex in two behaving monkeys engaged in a rapid face categorization task. Using multivariate pattern analysis (MVPA) techniques, we quantified at millisecond precision the face category information conveyed by IFPs in areas of the ventral stream. We further investigate the relationship between the selectivity and latency of individual electrodes as estimated with classical univariate vs. multivariate techniques and conclude on the similarity and differences between the two approaches.

1 Introduction

Object recognition in primates is mediated by the ventral visual pathway of the visual cortex. This pathway runs from the primary visual cortex (V1) through extrastriate visual areas II (V2) and IV (V4), to the inferotemporal cortex (IT) (which can be further subdivided in a posterior (PIT) and anterior (AIT) part), and then to the prefrontal cortex (PFC), which is involved in linking perception to memory and action. A widely held view is that recognition is achieved through computations that gradually increase both the selectivity and the invariance properties of the underlying visual representation at each successive stages of the ventral stream of the visual cortex [16].

Much of our understanding of the mechanisms underlying visual recognition has, however, focused on these areas studied individually. Previous electrophysiology and imaging studies in both humans and monkeys have emphasized the role of the ventral stream of the visual cortex in shape processing [13, 21] and, in particular, of higher areas such as the IT cortex [11, 18, 17] and the PFC [3, 4]

for the coding of abstract category information. Specifically, these studies have shown that category information can be decoded in these areas within times that are fast enough to be consistent with behavioral responses [7, 9, 10]. Relatively little is known about the contribution of lower and intermediate visual areas to the categorization process and the underlying dynamics in these areas.

Here we use a rapid visual presentation paradigm [19] to investigate the neural basis of face processing in the ventral stream of the monkey visual cortex. We recorded intracranial field potentials (IFPs) from intermediate areas of the ventral stream of the visual cortex (areas V2, V4 and PIT) from two monkeys while they were actively engaged in a rapid face categorization task. Using multivariate pattern analysis (MVPA) techniques, we quantified at millisecond precision the face category information conveyed by IFPs.

2 Methods

Two male rhesus macaques (M1 and M2, 14 year old) were chronically implanted with subdural macro-electrodes and performed a face vs. non-face categorization task using a go/no-go paradigm (12 / 16 electrodes in total for M1/ M2). Stimuli were presented very briefly at the center of the screen for 32 ms. A total of 80 images were organized in 4 visual subcategories (10 images each) for both targets (macaque, chimpanzee, human and other faces) and distractors (fruits, flowers, mugs, and car tires). The images were equalized in luminance and RMS contrast to prevent low level biases in the recordings. All training and experimentation procedures conformed to the French and European standards for the use of experimental animals; protocols were approved by the regional ethical committee.

Using EEGLab, data was down-sampled to 250 Hz and epoched (-50 ms – 200 ms around the each stimuli onset). Three epoch rejection criteria were also employed: voltage, i.e., the power for one electrode or across all electrodes exceeded 3 standard deviations from the mean value computed over the entire recording session. The latter criterion targeted motor artifacts. We also excluded any trials that contained a non-varying signal (i.e., zero variance) over a 10-ms time windows (i.e., due to amplifier saturation). The data was z-scored and weights across electrodes are thus interpretable.

Linear SVMs were trained for the classification of targets (faces) vs. distractors (objects) (C parameter optimized using a cross-validation procedure) for each monkey separately. Here we considered SVM classifiers trained directly on the average potential over each electrode computed over 10 ms-long sliding time windows. The training and testing splits included multiple repetitions of the stimuli collected over several days.

An initial analysis using training and test sets obtained from different recording days suggested that the neural signal was largely stable over the course of days (Figure 1). Based on these results, we pooled all recordings sessions together. Random splits were generated so that the identity of the stimuli did not overlap between training and test (all repetitions of a given stimulus were thus

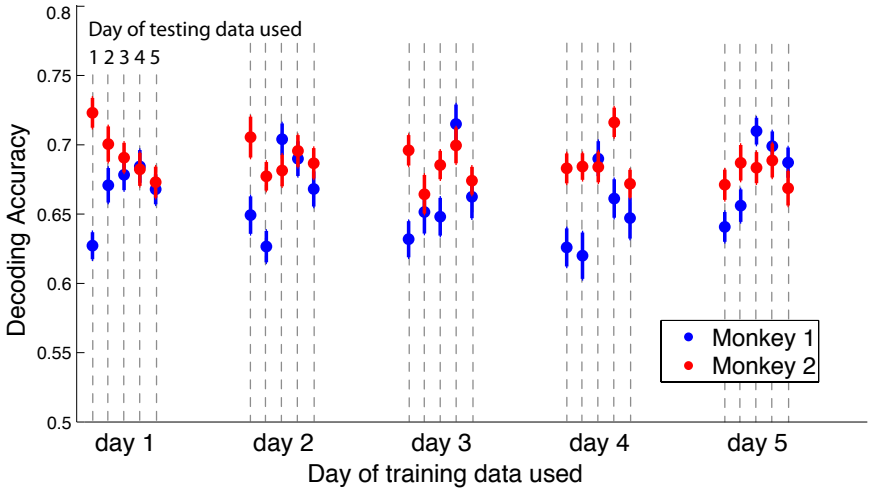

Fig. 1. Stability of the recorded signal for both monkeys: Decoding accuracy is reported for every possible training-testing set combination for different days.

used for either training or testing). This limits the possibility that the classifier simply "memorizes" patterns of neural activity associated with individual stimuli and instead pick on high-level category information.

Our analysis focuses on the initial 200 ms after stimulus onset to rule out any possible contamination from motor activity. We report the mean classification accuracy estimated over 50 cross-validations where the data was randomly split into two mutually exclusive halves for training and test (95% confidence intervals were computed based on the activity during the baseline period [-50ms 20ms]. These are indicated as vertical dashed lines around the chance level of decoding accuracy in Figures 2C and 4B). Two conditions are assumed to exhibit a statistically significant difference if the corresponding classifiers performance fall outside of these confidence intervals. This basic procedure was employed for all of decoding results, although the variables (i.e., electrodes included in the decoding) and the choice of decoding time points were manipulated for different types of analyses. Unless otherwise noted, classifiers are trained and tested on the same time points in the epoched and smoothed voltage data.

3 Results

The spatio-temporal nature of visual processing: The results of the neural decoding analysis are reported in Figure 2. Shown are the estimated latencies such that the decoding performance deviates significantly from baseline (60 ms for M1 and 88 ms for M2; see Methods). Overlaid on Figure 2A are the event related potentials (ERPs) for single electrodes. These ERPs, highlight both functional similarities within and differences between the three groups of electrodes (V2, V4

and PIT) thus confirming our initial anatomical localization. Figure 2B shows the corresponding weights of the linear SVM classifier averaged over each random split for every time point and expressed as a percentage of the maximum weight. These weights highlight the contribution of each electrode at specific time points and suggest that, at each time point, different areas and electrodes seem to encode the relevant category information. For M1, the contribution of electrodes over time seems to follow the known visual hierarchy (V2 $\rightarrow$ V4 $\rightarrow$ PIT).

Could the rapid decoding of category information from distributed patterns of neural data be predicted from univariate data analysis? Figure 2C shows a comparison between this multivariate analysis and similar analysis conducted on single electrodes (using the same classification procedure). The performance of a classifier trained on multiple electrodes distributed over areas V2, V4 and PIT exceeds the performance of the best electrode for every time point. This suggests that most of the time (for both monkeys, but more clearly for M2), electrodes in isolation do not provide nearly as much information as their combination. Overall this analysis reveals a very rapid signature for high-level category within 100 ms (in agreement with human electrophysiology recordings [10, 20]), which is supported by a distributed representation over areas and electrodes and that evolves over time as the signal progresses through the visual hierarchy.

Univariate vs. multivariate pattern analysis: Figure 3A provides a comparison between classical univariate significance tests for each electrode, the time points with the corresponding weights acquired from the MVPA analysis, and the corresponding univariate decoding performance for this electrode using the same SVM classifier ($p < 0.01$ for all tests; correction for multiple comparisons is approximated by seeking consistent significance for at least ten consecutive time

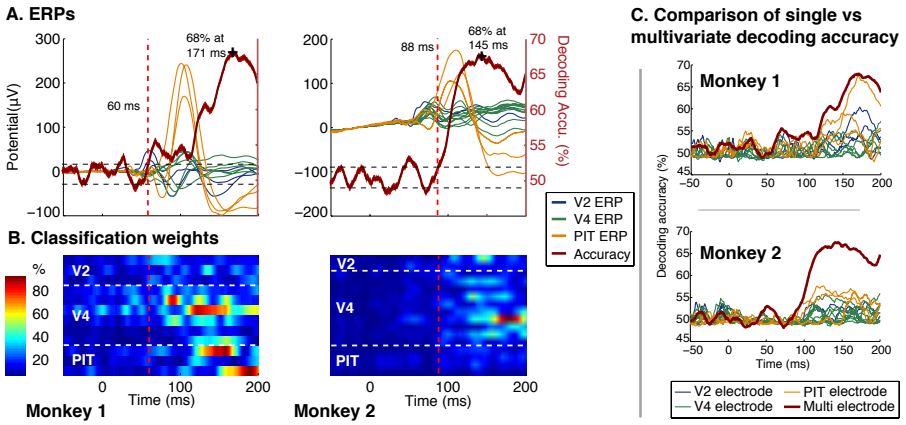

Fig. 2. A) Overall classification accuracy as a function of time; B) corresponding classification weights for each electrode and time point and C) decoding accuracy for single vs. multiple electrodes.

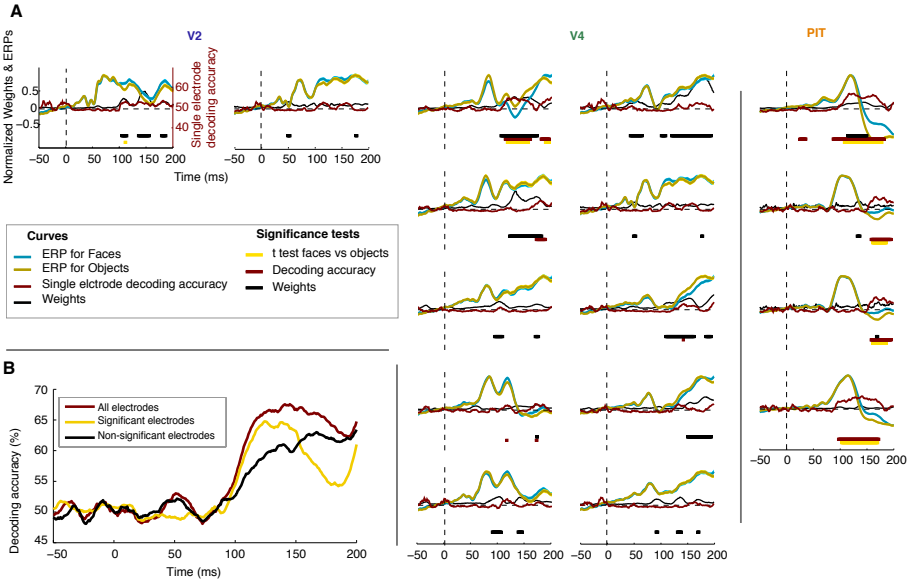

Fig. 3. Single versus multiple electrode analysis of informativeness measures for M2. A) ERP comparison between targets (cyan) and distractors (yellow) with decoding performance for the corresponding electrode (red) and weights resulting from MVPA (black). T-tests for targets vs. distractors are shown with bars. B) MVPA decoding performance using all vs. subsets of electrodes with significant (yellow lines) & non-significant differential activity.

points. This is a common practice employed by other researchers [19] and it also seems to be compatible with Bonferroni and FDR corrections. We chose this method for a more intuitive explanation). Interestingly, both t-test and univariate decoding exhibit a very similar trend. However, the electrode contributions assessed via MVPA seems to indicate qualitative diffxerences. For instance, for both monkeys, some electrodes (mainly in V4) seem to be associated with high classification weights although these electrodes do not seem to exhibit any significant differential activity as measured with classical univariate analysis. Overall this suggests that classical univariate analyses may underestimate the contributions of single electrodes. Last, Figure 3B shows that removing electrodes that exhibit significant differential activity tend to slow down the decoding of category information.

Characterizing the nature of the neural code We also conducted a *discriminant cross-training* (DCT) analysis [12]: A classifier is trained using patterns of neural activity at one time point t_o and tested on all time points in the epoch, producing a matrix of decoding accuracy for each combination of training/testing time points (figure 4A). Note that the diagonals of these two matrices correspond to the decoding accuracy curves shown in figure 2C. This analysis allows us to

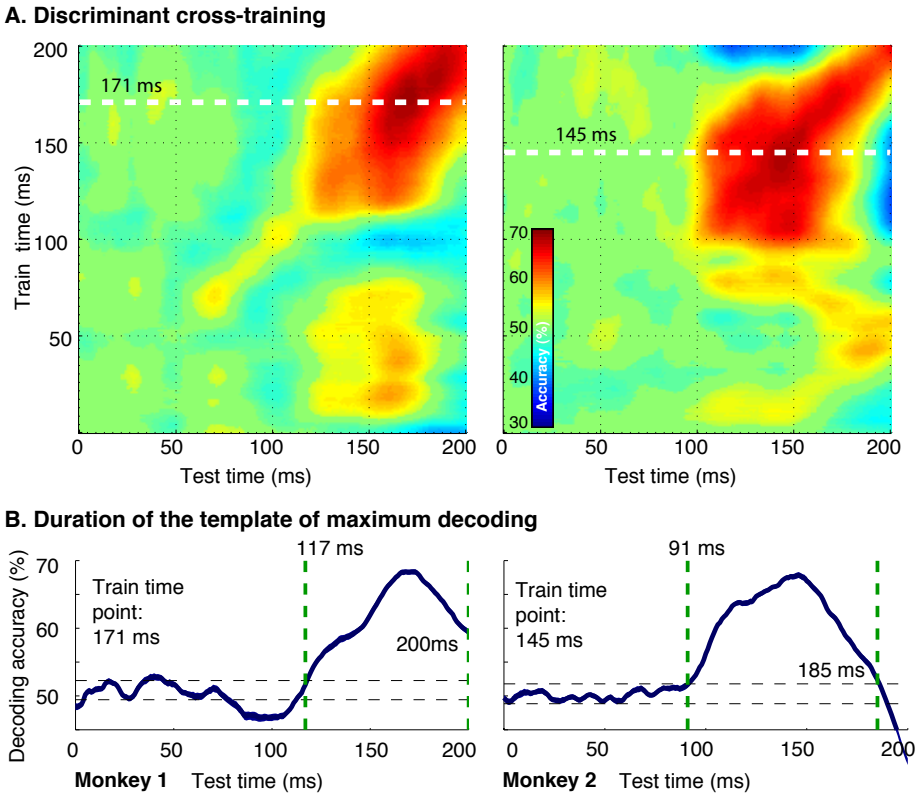

Fig. 4. A) Discriminant cross-training (DCT) analysis: Color indicates decoding accuracy for each combination of training and test times (vertical and horizontal axes respectively). B) Decoding accuracy for a classifier trained on the time point where the overall decoding was maximal (dashed line, see figure 2), and tested on the remaining time points.

assess how suitable is the neural template picked up by the classifier over time. As seen on figure 4A, the neural template that emerges around 100 ms post-stimulus (X axis), and remains stable for about 100 ms, suggesting a sustained coding of facial information in this set of areas possibly supported by feedback. Figure 4B provides a cross-section taken from the upper graphs (on dashed white lines). The rationale behind taking these points in time to show the duration of the "neural template" was the assumption that category information would be maximal at these time points in the trial and therefore they would be the best approximations of temporally adjacent task relevant brain states.

4 Discussion

Psychophysical studies using the rapid categorization paradigm [19], have demonstrated the remarkable speed at which the visual system can operate (with reaction times as fast as 200 ms for monkeys and 250 ms for human participants [2]). The present study provides a quantitative description of the signal propagated through the ventral stream of the visual cortex during the task. The present analysis suggests latencies for the coding of face category that are shorter than previously estimated from single electrode IT recordings [8]. The very short latencies reported in the present study are, however, compatible with those estimated from human electrophysiology [10, 20] and psychophysics experiments [1, 6]. These results may help shed some light on a long-standing debate on the role played by various EEG/MEG components in face perception. However it is important to emphasize that care should be taken in interpreting these latencies as some of these differences may be influenced by the task difficulty and the stimulus set used. The present decoding analysis may highlight a monkey homologue of the P100 assumed to reveal one of the first stages of face processing [5, 15] starting in the Occipital Face Area (OFA).

The present study also demonstrates the stability of the neural signal over days of recording and the ability of a linear classifier to genralize over recording session. Furthermore our analysis highlights some of the key benefits of MVPA as compared to a classical ERP/univariate analysis by suggesting that intermediate visual areas may contribute relatively early coding signals for the processing of faces. This idea seems compatible with the high level of performance obtained with low-level visual features for the processing of faces [14]. Overall this suggests that intermediate visual areas such as V4 may already play a key role in visual recognition and that the coding of category information may be implemented via a distributed process in both space and time.

5 Acknowledgements

The data collection part of this work was supported by the Centre National de la Recherche Scientifique, Université de Toulouse, Université Paul Sabatier and the Fondation pour la Recherche Médicale. The Délegation Générale pour l'Armement provided financial support to MC. The data analysis part of this work was supported by DARPA grant to TS. Additional support was provided by Brown University, Center for Computation and Visualization, and the Robert J. and Nancy D. Carney Fund for Scientific Innovation.

References

1. Crouzet, S.M., Kircher, H., Thorpe, S.J.: Fast saccades toward faces : Face detection in just 100 ms. Journal of Vision 10(4), 1–17 (2010)
2. Fize, D., Cauchoix, M., Fabre-Thorpe, M.: Humans and monkeys share visual representations. Proceedings of the National Academy of Sciences of the United States of America 108(18), 7635–40 (May 2011)

3. Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K.: Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291(5502), 312–6 (2001)
4. Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K.: A comparison of primate prefrontal and inferior temporal cortices during visual categorization. Journal of Neuroscience 23(12), 5235–46 (2003)
5. Freiwald, W.A., Tsao, D.Y.: Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330(6005), 845–51 (2010)
6. Girard, P., Jouffrais, C., Kirchner, C.H.: Ultra-rapid categorisation in non-human primates. Animal Cognition 11(3), 485–93 (2008)
7. Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J.: Fast readout of object identity from macaque inferior temporal cortex. Science 310(5749), 863–6 (2005)
8. Kiani, R., Esteky, H., Tanaka, K.: Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. Journal of Neurophysiology 94(2), 1587–96 (2005)
9. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A.: Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60(6), 1126–41 (2008)
10. Liu, H., Madsen, J.R., Agam, Y., Kreiman, G.: Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62(2), 281–290 (2009)
11. Logothetis, N.K., Sheinberg, D.L.: Visual object recognition. Annual Review of Neuroscience 19(1), 577–621 (1996)
12. Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., Poggio, T.: Dynamic population coding of category information in inferior temporal and prefrontal cortex. Journal of Neurophysiology 100(3), 1407–19 (Sep 2008)
13. Pasupathy, A., Connor, C.E.: Population coding of shape in area V4. Nature Neuroscience 5(12), 1332–8 (2002)
14. Pinto, N., DiCarlo, J.: How far can you get with a modern face recognition test set using only simple features? Vision and Pattern Recognition, pp. 2591–2598 (2009)
15. Pitcher, D., Walsh, V., Duchaine, B.: The role of the occipital face area in the cortical face perception network. Experimental Brain Research 209(4), 481–93 (2011)
16. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature Neuroscience 2(11), 1019–25 (1999)
17. Sigala, N., Logothetis, N.K.: Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415(6869), 318–20 (2002)
18. Tanaka, K.: Inferotemporal cortex and object recognition. Annual Review of Neuroscience 19, 109–139 (1996)
19. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. Letters to Nature 381, 520–522 (1996)
20. Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M.A., Adolphs, R.: Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. PloS One 3(12), e3892 (2008)
21. Ungerleider, L.G., Bell, A.H.: Uncovering the visual "alphabet": advances in our understanding of object perception. Vision Research 51(7), 782–99 (2011)

3.4 Codage neuronal des objets et catégories chez le singe

3.4.1 Introduction

Où et comment se forment les représentations visuelles qui nous permettent de comprendre le monde qui nous entoure? En particulier, quelles sont les structures, les réseaux et les mécanismes qui nous permettent de catégoriser ou de reconnaître les objets?

Qu'est-ce que la reconnaissance d'objets ?

La reconnaissance d'objet peut être définie comme la capacité qui permet d'assigner un label (nom, réponse comportementale) à un objet particulier. Ce label peut être précis: on parle alors d'identification, ou général : on parlera de catégorisation (DiCarlo et al. 2012). La reconnaissance d'objet est un synonyme d'abstraction visuelle puisque, pour réaliser cette tâche, il faut faire abstraction de toutes les variations fortuites ou accidentelles (en taille, en position, en luminance, d'orientation et de contexte) que présente chaque image d'un même objet, reçue sur la rétine.

Sélectivité et invariance

Pour qu'un système neuronal soit impliqué dans la reconnaissance d'objets il faut qu'il soit à la fois :

- *Sélectif* : coder un nombre limité d'objets ou catégories.
- Et *Invariant (ou tolérant) :* ne pas être affecté par les différentes variations que les objets peuvent présenter dans le monde réel

Expérimentalement, on évalue la sélectivité en présentant un grand nombre stimuli différents et en montrant la spécificité des réponses neuronales pour une catégorie ou un objet en particulier. L'invariance peut être testée de deux manières :

- soit en utilisant des stimuli bien contrôlés, présentés de manière isolée sur un fond gris. Dans ce cas on peut faire varier de manière discrète une ou plusieurs propriétés accidentelles de l'objet (taille, position..) et montrer l'invariance des réponses neuronales face à ces variations (Hung et al. 2005).

- Soit en utilisant une large banque de photographies naturelles des objets, qui présentent de manière inhérente une combinaison continue de toutes les propriétés accidentelles possibles (Thorpe et al. 1996)

Dans les deux cas le système neuronal testé sera invariant s'il répond de manière similaire à toutes les variations de l'objet ou de la catégorie pour lesquels il est sélectif.

La voie visuelle ventrale substrat de la reconnaissance d'objets ?

Chez le singe, après s'être fait la main sur le cortex visuel primaire, les premiers électrophysiologistes se sont mis à explorer d'autres régions du cerveau. En enregistrant dans le lobe temporal, ils se sont rendus compte que des neurones répondaient sélectivement à certains stimuli complexes et notamment aux visages (Gross et al. 1972, Desimone et al 1984). Les études de lésions spécifiques de la voie visuelle ventrale ont souligné son rôle crucial dans la discrimination visuelle et la reconnaissance d'objet (Schiller 1995) ; un rôle confirmé par des expériences de microstimulations récentes. Afraz et al. (2006) en stimulant certaines populations de neurones dans le cortex inférotemporal (IT) montrent un lien de causalité fonctionnelle entre l'activité de ces neurones et la catégorisation des visages.

Chez l'humain, des lésions corticales pathologiques le long de l'axe occipito-temporal, provoquent aussi des déficits de reconnaissance visuelle pour certains objets (Farah 1992). De même la technique de stimulation transcrânienne magnétique, adaptée à l'homme, révèle une causalité fonctionnelle entre ces régions et la perception des objets (Pitcher et al. 2009).

Organisation hiérarchique de la voie visuelle ventrale

La voie visuelle ventrale est composée de différentes aires visuelles interconnectées. Chaque aire est définie selon des critères cytoarchitectoniques et/ou fonctionnels (Felleman et Van Essen 1991). Des cartes rétinotopiques complètes ont été retrouvées pour V1, V2 et V4 (Felleman et Van Essen 1991), chacune de ces aires transforme la totalité de la scène visuelle en une représentation neuronale. Pour IT, une rétinotopie grossière existe dans la partie postérieure (Boussaoud et al. 1991) mais pas pour les régions centrales et antérieures (Fellman et Van Essen 1991). C'est pourquoi même si on a souvent divisé IT en sous-régions, comme TEO et TE (Janssen et al. 2000) ou,

comme dans la Figure 10, en postérieur (PIT), central (CIT) et antérieur (AIT) (Felleman et Van Essen 1991), il reste encore des doutes sur le fait qu'IT constitue une ou plusieurs aires visuelles. Des études récentes en IRMf chez le singe montrant trois (Tsao et al. 2003), six (Tsao et al. 2008) ou plus (Ku et al. 2011) petits regroupement de cellules, impliquées dans le traitement des visages, semblent indiquer que l'arrangement spatial des neurones au sein du cortex inférotemporal dépendrait plus des buts comportementaux (ie : catégorisation ou identification d'objets) qu'à une organisation rétinotopique. Un principe qui semble aussi s'appliquer chez l'homme (voir plus loin).

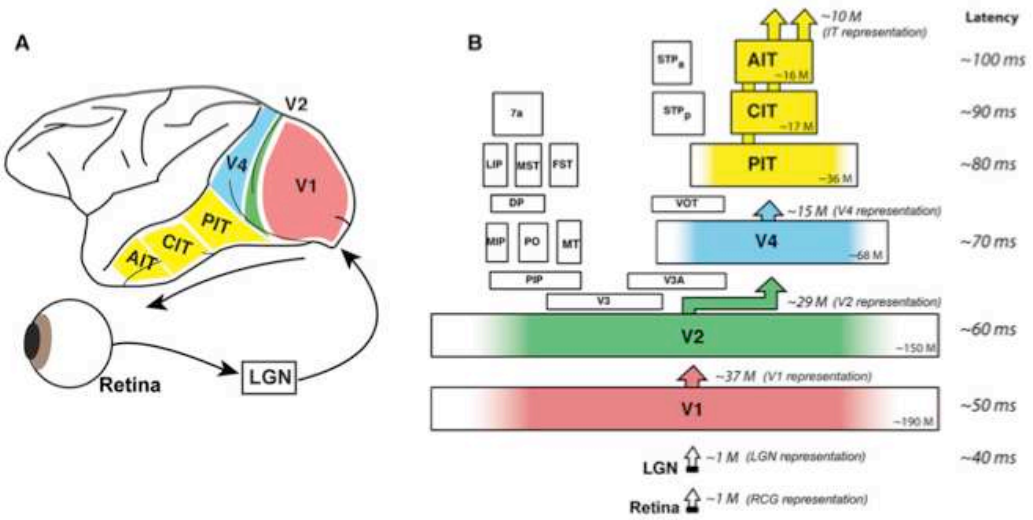

Figure 10 : Représentation schématique de l'organisation de la voie visuelle ventrale chez le singe (d'aprés Dicarlo et al. 2012). (A) Les différentes aires visuelles depuis la rétine jusqu'au cortex inféro-temporal. (B) Chaque aire est représentée par des rectangles dont la taille est proportionnelle à sa surface corticale (selon les travaux de Felleman and Van Essen, 1991). Le nombre approximatif de neurones présents dans les 2 hémisphères cérébraux est présenté dans le coin droit de chaque rectangle (en millions) et le nombre de neurones de projection est présenté sous chaque rectangle (selon les travaux de Collins et al. 2010). Enfin les latences médianes des réponses neuronales sont présentées à droite (d'après Bullier 2001).

Toutes les aires corticales visuelles sont constituées de 6 couches de neurones. L'information entre les différentes aires circule de manière bien définie. Le signal ascendant (feedforward) entrant est reçu dans la couche 4 et transmis par les couches supérieures aux aires plus antérieures. Le signal descendant (feedback) est produit par les couches inférieures et reçu dans les couches supérieures et inférieures des aires plus postérieures (Felleman et Van Essen 1991). Cette organisation des aires et la circulation de l'information dans l'axe postéro-anterieur (occipital-temporal) sont en

faveur d'une organisation hiérarchique de la voie visuelle ventrale. D'abord l'information visuelle est projetée sur la rétine, puis transmise au corps genouillé latéral (LGN) du thalamus puis dans le cortex, où elle progresse de V1 à IT en passant par V2 et V4 (Figure 10, Felleman et Van Essen 1991). Ces résultats sont confortés par les latences d'activation des neurones dans ces différentes aires, décalées d'environ 10 ms par aire en progressant dans la hiérarchie (Figure 10, Bullier 2001), et les traitements visuels de complexités croissantes depuis V1 jusqu'à IT (Kobatake et Tanaka 1994a, Serre et al. 2007, Marr 1982). Ainsi environ 100 ms après que l'image a été projetée sur la rétine, une première vague de sélectivité neuronale est observée dans IT (Desimone et al. 1984, Hung el 2005, Kobatake et Tanaka 1994a, DiCarlo et al. 2012, Logothetis et Sheinbeirg 1996).

Hypothèses sur les algorithmes permettant le codage des objets dans la voie visuelle ventrale

On est encore très loin de comprendre comment le cerveau transforme l'entrée rétinienne en une collection d'objets pleins de sens telle qu'on la ressent. Néanmoins, en se basant sur l'organisation hiérarchique des aires visuelles et les nombreuses connections ascendantes et descendantes, deux hypothèses sur les mécanismes permettant de construire une représentation visuelle élaborée ont été proposées.

La première est celle du *codage ascendant seul* (« pure feedforward ») dans lequel chaque aire visuelle successive, en grimpant la hiérarchie apporte un pouvoir de calcul supérieur à la précédente et permet de résoudre des tâches de complexité croissante (Marr 1982, Riesenhuber et Poggio 1999, Serre et al. 2007, Dicarlo et al. 2012). Une analogie assez explicite de cette hypothèse est celle du montage en chaîne d'une voiture. A travers les différentes stations de travail d'une usine automobile, en partant de la matière première, des pièces simples sont élaborées puis associées pour former des structures de plus en plus complexes jusqu'à l'aboutissement final de la voiture entière.

L'hypothèse alternative, *de codage par interaction ascendante et descendante*, postule que les interactions entre les différentes couches par des boucles dîtes récurrentes (information transmise de l'aire A1 à l'aire A2 par des connexions ascendantes, puis retour sur A2 par des connexions descendantes) seraient un algorithme clé de la construction du percept visuel (Mumford 1991, Bullier 2001,

Roelfsema et Houtkamp 2011). Ici l'analogie pertinente est celle d'une armée où les fantassins (neurone de V1) transmettent une information incertaine (« je vois peut-être un contour ») au sergent (V2), qui passe l'ensemble des informations qu'il a reçues au lieutenant (V4) et ainsi de suite. Les généraux à la tête de l'armée (IT, préfrontal Cortex) obtiennent une idée grossière de la scène visuelle (« the gist ») qui permet par des processus descendants de préciser le travail local des fantassins (Bar et al. 2006, Hochstein et Ahissar 2002).

3.4.2 Article 3 : Des représentations précoces dans les aires visuelles intermédiaires permettent la catégorisation rapide

Problématique de l'article

Quel est le point commun entre une baleine, un moustique, un raton laveur et un pigeon? Visuellement ils sont tous extrêmement différents; regrouper ces différents objets en une seule et même catégorie, celle d'animal, semble être une tâche nécessitant un très fort degré d'abstraction visuelle. Comment, où et en combien de temps le cerveau des primates peut-il former un concept aussi abstrait reste une question non résolue. On trouve chez l'homme des neurones répondant spécifiquement à la catégorie animale, dans des aires d'intégration haut niveau comme l'hippocampe (Kreiman et al. 2000) ou l'amygdale droite (Mormann et al. 2011). Néanmoins, la latence de réponses de ces neurones (environ 320 ms) semble difficilement compatible avec la rapidité à laquelle on peut détecter ou reconnaître un animal dans une scène naturelle.

Effectivement, comme nous l'avons vu dans le chapitre 2, dans un protocole de réponse manuelle, les singes et les hommes sont capables de catégoriser des images naturelles comme contenant ou non un animal en une fraction de seconde (Thorpe et al. 1996, Fabre Thorpe et al. 1998). En mettant au point une tâche de choix forcé saccadique, où deux images sont flashées simultanément, Kirchner et Thorpe ont même montré que les humains étaient capables de saccader vers l'image contenant un animal avec des latences aussi courtes que 150 ms (Kirchner et Thorpe 2005). Des macaques rhésus peuvent effectuer cette tâche en 100 ms (Girard et al. 2008) et des humains à qui l'on demande de saccader vers des visages en présentant des véhicules dans l'autre hémichamp peuvent le faire en 100 ms (Crouzet et al. 2010). De telles latences comportementales laissent peu de temps pour un traitement neuronal élaboré des stimuli visuels. Elles excluent, de toute évidence, l'implication des neurones "animaux" du lobe temporal médian et sont difficilement compatibles avec les latences d'activité catégorielle reportée dans les aires frontales (PFC 160 ms) et pariétales (LIP 130 ms, Swaminathan et Freedman 2012).

Une piste pour élucider comment le cerveau peut accéder à une abstraction visuelle haut-niveau en un temps de calcul limité, nous vient du domaine de l'intelligence

artificielle et en particulier de celui de la vision par ordinateur. Effectivement plusieurs modèles de vision par ordinateur ont montré que des traits visuels simples (Torralba et Oliva 2002) ou de complexité intermédiaire (Ullman 2002, Serre et al. 2007) étaient suffisants pour former des catégories d'objets même au niveau abstrait de la catégorie animale. Les neurones permettant de coder ce genre de traits visuels sont présents tôt dans la hiérarchie du système visuel ventral (V4/PIT, Pasupathy 2006) et peuvent répondre avec des latences de potentiel d'action aussi courtes que 60-70 ms (Bullier 2001).

Hypothèse

Dans cet article, on pose l'hypothèse de la capacité de la voie visuelle ventrale à extraire rapidement des traits visuels de complexité intermédiaire pour résoudre une tâche nécessitant un fort degré d'abstraction visuelle en un temps minimum.

Méthodes

Pour cela deux macaques rhésus, entraînés à faire la tâche de catégorisation rapide animal - non animal, on été équipés avec des macroélectrodes pour l'enregistrement de potentiels de champs intracrâniens dans la voie visuelle ventrale. En utilisant la méthode de décodage des potentiels de champs décrite dans le chapitre précédent, on caractérise et quantifie les représentations mises en jeu dans la voie visuelle ventrale pour résoudre cette tâche de catégorisation abstraite. Mieux, ces méthodes nous permettent de lier traitement visuel et comportement.

Résumé des résultats

On montre dans cet article que :

- on est capable de décoder, essai par essai, des représentions correspondant à une *catégorie superordonnée*, dans les aires visuelles intermédiaires (PIT,V4) du singe.
- ce signal catégoriel émerge *rapidement* après la présentation du stimulus (moins de 100 ms)
- tout comme le comportement, le décodage précoce de la catégorie *n'est pas affecté par la présention d'un masque rétrograde* 50 ms après la présentation du stimulus
- ces représentations catégorielles sont *utilisées par le singe* pour résoudre la tâche. La latence de décodage prédit la latence des temps de réaction et la précision de décodage neuronal corrèle avec la précision de catégorisation des singes.
- ces représentations concordent avec le fonctionnement d'un *modèle hiérarchique ascendant du système visuel*.
- l'activité neuronale décodée ici n'est pas purement liée à la nature cible ou distracteur des stimuli mais en représente bien le *contenu visuel*. Au sein de ses cibles, on est capable de décoder significativement des sous-catégories animales fines.
- Ces représentations reflètent la *nature conceptuelle* de la capacité de catégorisation des singes. Tout comme les animaux, un classifieur entraîné à catégoriser des images familières est capable de *généraliser* à de nouveaux stimuli.

ARTICLE 3: Early ventral stream visual activity enables rapid visual categorization

Maxime Cauchoix, Sebastien Crouzet, Denis Fize, Thomas Serre

En préparation

Early ventral stream neural activity enables rapid visual categorization

Maxime Cauchoix [† 1,2], Sébastien M. Crouzet [† 3], Denis Fize [1,2] & Thomas Serre [3]

[1] Centre de Recherche Cerveau et Cognition, Université Paul Sabatier, Université de Toulouse, Toulouse, France
[2] Faculté de Médecine de Purpan, CNRS, UMR 5549, Toulouse, France
[3] Cognitive, Linguistic and Psychological Sciences Department, Institute for Brain Sciences, Brown University, Providence, RI 02912, USA
[†] These authors contributed equally to the work

Primates can recognize objects embedded in complex natural visual scenes in a glimpse. Rapid categorization paradigms have been extensively used to study the psychological basis of high-level object recognition [1-3]. However, their neural underpinning remains to be understood, and the incredible speed of sight has yet to be reconciled with modern ventral stream cortical theories of shape processing [4].

Here we recorded multi-electrode intracranial field potentials (IFPs) from intermediate areas (V4/PIT) of the ventral stream of the visual cortex while monkeys were actively engaged in detecting the presence or absence of animal targets in natural scenes. Reliable neural decoding of superordinate category information was possible shortly after stimulus onset from single trials. This early selective signal was unaffected by the presentation of a backward (pattern) mask and was tightly linked to accuracy and reaction time. Together, the present study establishes that early feedforward neural activity in the ventral stream induces a selective signal subsequently used to drive rapid natural scene categorization.

Results

Two monkeys (M1 and M2) were trained to perform a rapid categorization task [5] (Fig. 1) in which they had to report the presence of an animal in briefly presented natural scenes (Fig. S1). After training using a familiar set of images, both monkeys were able to detect the presence of an animal with a high level of accuracy (M1=95%; M2=93%).

Intracranial Field Potentials (IFPs) were recorded from multiple electrodes implanted along the ventral stream of the visual cortex in both monkeys (Fig. S2). We conducted a (temporal) multi-variate pattern analysis by training and testing a linear classifier (SVM) directly on the amplitude of the neural signal (downsampled at 512 Hz) distributed across all electrodes (M1: 9 electrodes; M2: 13 electrodes), independently for every time point. This analysis demonstrates that superordinate category information (animal vs. non animal) can be reliably decoded from single trials very early (~90-100 ms post stimulus onset; Fig. 2a).

On half the trials, a backward mask (1/f pink noise) was presented following stimulus presentation (50 ms post stimulus onset). This type of mask reduces the stimulus visibility and was used in previous psychophysics studies to try to limit contributions from cortical feedback during visual processing [6, 7]. The overall behavioral accuracy (M1: 93%; M2: 90%) was significantly reduced on masked trials (M1: 87%, $\chi^2(1)=99$, $p<0.001$; M2: 86%, $\chi^2(1)=40$, $p<0.001$) but far exceeded chance level (M1: $\chi^2(1)=2059$, $p<0.001$; M2: $\chi^2(1)=2326$, $p<0.001$).

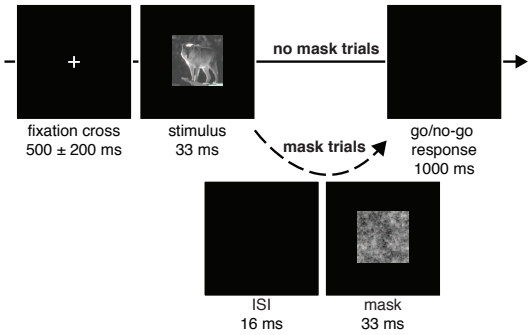

Figure 1. Go/no-go scene categorization task. Following the presentation of a fixation cross, an image is flashed for 33 ms. Monkeys indicated whether an animal stimulus was present by releasing a button within 1 s following image presentation. On half the trials, a backward (pattern) mask (1/f pink noise) was displayed for 33 ms following a 16-ms blank screen (stimulus onset asynchrony, SOA = 50 ms).

Cumulative d' curves plotted as a function of response times are shown in Fig. 2b. Cumulative number of hit and false alarm responses were used to calculate z(Hits) - z(False Alarms) at each time point (where z is the inverse of the normal distribution function). These types of discriminability curves provide a behavioral estimate of the processing dynamics. The shortest minimal reaction time (see Methods), which corresponds to the shortest "reliable" behavioral responses and is typically associated with feedforward processing [8], was unaffected by the presentation of the mask. A significant difference between masked and unmasked trials was only apparent for slower trials (with a delay ~50 ms; Fig. 2b).

MVPA revealed a consistent pattern of results (Fig. 2a; see also refs [9-11]), whereby early selective neural activity (starting ~90-100 ms

post stimulus onset) remains unaffected by the presentation of a mask, which seems to only interfere with visual processing during a later time window (~50 ms later). These results are thus consistent with masking theories positing that visual processing of the stimulus and mask are kept separate during an initial short time period and that these initial vs. late time windows roughly correspond to feedforward and recurrent processing. In addition, the effect of the mask was delayed by ~50 ms for both the neural and behavioral responses, a hallmark of feedforward processing in a pipeline-like architecture [8, 12]. In subsequent MVPA analyses, we thus focused on this early ventral stream neural activity (<140 ms after stimulus onset) and pooled masked and unmasked trials which are indiscriminable during this time interval.

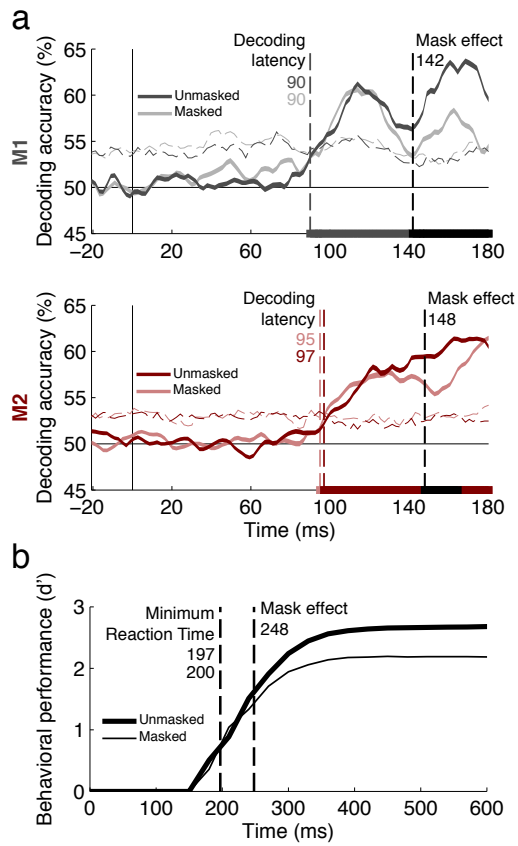

Figure 2. Linking early ventral stream neural activity with feedforward processing. a) Decoding accuracy from ventral stream neural activity (curves correspond to the non-parametric 95% confidence intervals of the mean) for masked (light curves) vs. unmasked (dark curves) stimulus presentations for monkey M1 (gray) and M2 (red). Dotted vertical lines indicate decoding latencies for the masked and unmasked conditions and the earliest significant mask effect (see also horizontal bars at the bottom of the graph). Dotted curves indicate chance level (upper limit of the 95% confidence interval). b) Cumulative d' curves plotted as a function of response times. Minimal reaction times and earliest significant mask effects shown as vertical lines.

Does early ventral stream neural activity influence subsequent animal behavior? To link this early ventral stream neural activity to behavioral accuracy, we computed a discriminability score for each individual image based on the decoded neural activity and monkeys' behavioral responses separately. These "animalness" scores reflect the likelihood for an image to be classified as an animal irrespective of its actual label (see Fig. S3 and Experimental procedures). Such score provides a compact characterization of the visual strategy employed by a visual system and permits the comparison between monkey neural and behavioral data via direct correlation between these scores. The correlation between animalness scores computed from neural activity and behavioral responses was highly significant (M1: r^2=0.45, p<0.001; M2: r^2=0.27, p<0.001), even after correcting for category using partial correlation (see Methods; M1: r^2*=0.29, p<0.001; M2: r^2*=0.15, p<0.001).

We computed a similar score for a representative feedforward computational model of the ventral stream of the visual cortex (Hmax) which was previously shown to match human performance on the same rapid animal categorization task [7]. A significant correlation was also found between the Hmax model and monkey behavior (M1: r^2=0.41, r^2*=0.16, p<0.001; M2: r^2=0.41, r^2*=0.13, p<0.001) as well as neural activity (M1: r^2=0.28, r^2*=0.14, p<0.001; M2: r^2=0.20, r^2*=0.09, p<0.001). In contrast, no significant correlation was observed between neural activity and a low-level visual representation based on pixel intensities (M1: r^2≃0, r^2*≃0, p=0.8; M2: r^2=0.02, r^2*=0.02, p=0.06).

To test for a link between this early neural activity and behavioral response times, a hallmark of decision making processes [13, 14], we organized trials into quartiles based on the overall distribution of reaction times (Fig. 3a). MVPA (over a 90-140 ms post-stimulus onset time window) revealed a monotonic (near-linear) relationship between decoding accuracy and mean reaction times for each quartile (Fig. 3a): Trials corresponding to faster quartiles were decoded with higher accuracy (M1: F(3,1196)=557, p<0.001; M2: F(3,1196)=408, p<0.001). A temporal MVPA approach conducted on individual quartiles separately (Fig. 3b) further suggested that decoding latencies very closely followed response times. Trials associated with faster quartiles were decoded faster and with higher accuracy. Overall, these results demonstrate that early feedforward neural activity in the ventral stream is tightly linked to both

behavioral accuracy and reaction time.

We further tested the selectivity of this early ventral stream neural activity for fine level visual information. Reliable decoding of target subcategories (people, macaques, chimpanzees, otter faces; Fig. 4a) from target images suggests that top-down task-related information does not completely override the visual information contained in early visual processes and that, in principle, early ventral stream neural activity could subserve multiple visual recognition tasks [15, 16].

familiar (Fig. 4b; M1: 65%, $p<0.01$; M2: 66% $p<0.01$) to the novel set (M1: 60%, $p<0.05$; M2: 57%, $p<0.05$).

Discussion

Evidence suggesting that monkeys are capable of learning a high-level visual concept of abstract natural categories has so far been limited to behavioral studies [5, 17]. Previous monkey electrophysiology studies on category learning (e.g., cats vs. dogs [18] or faces vs. fishes [19]) did not test for generalization to

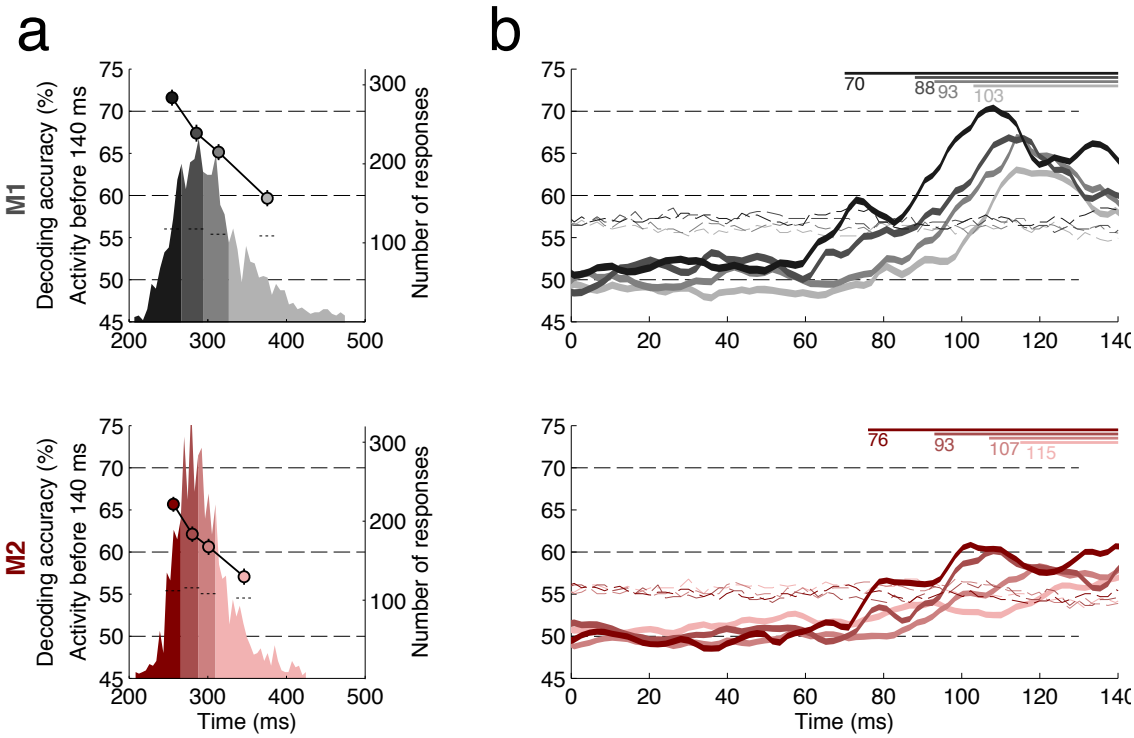

Figure 3. Linking early ventral stream neural activity with response times. a) Trial binning according to monkeys reaction times and corresponding neural decoding (overlaid on the distributions) shown as circles (90-140 ms time window; 95% confidence interval shown as error bars) for monkey M1 in gray and M2 in red. b) MVPA conducted on individual quartiles. Significant deviation from chance level is shown at the top of the graph with horizontal bars. Horizontal dotted lines / curves indicate chance level (upper limit of the 95% confidence interval).

To further verify that monkeys did form a high-level concept of animal category beyond simply memorizing individual images, we evaluated behavioral responses and neural decoding accuracy separately for a novel and familiar set of images (see Methods). On the very first presentation of the novel stimuli (80 images × 5 days), the accuracy of the monkeys remained well above chance (M1: 77% accuracy, $\chi^2(1)=96$, $p<0.001$; M2: 82%, $\chi^2(1)=123$, $p<0.001$). This level of accuracy was directly comparable to untrained human participants on the same set of stimuli [7] (n=21, 83% accuracy). Similarly, decoding from early ventral stream neural activity (90-140 ms time window post-stimulus onset) generalized well above chance from the

novel (unfamiliar) stimuli and are thus compatible with rote learning of individual exemplars. The present study establishes a link between ventral stream neural activity and primates' generalization ability during high-level processing of natural scenes.

A widely held view is that primates' "core object recognition" system proceeds through a cascade of hierarchically organized areas along the ventral stream of the visual cortex with computations at each successive stage being largely feedforward [15, 20]. However, surprisingly, evidence for this "feedforward hypothesis" has remained, at best, indirect.

The hypothesis is supported most directly by the short time spans required for category information to be readout from temporal lobe

areas in human [21] and non-human primates [22, 23]. Nonetheless, a link between this initial feedforward sweep of activity and behavioral responses has yet to be demonstrated. In particular, it has been shown that feedforward activity is not sufficient and that recurrent activity is necessary for conscious visual perception [24].

Here, using multi-electrode recording techniques in intermediate areas (V4/PIT) of the ventral stream, we establish a link between early feedforward ventral stream neural activity and behavioral responses (both accuracy and reaction time) during rapid visual categorization in natural scenes. In contrast to the relatively late (≈320 ms) animal category selective responses previously observed in the human amygdala [25], the present study establishes reliable animal category information can be readout very early (<100 ms) from visual cortex and that this early selective signal is tightly linked to accuracy and reaction time. This latency also overlaps with the optimal timing for microstimulations in IT to affect perceptual decisions [26].

Most previous attempts to link neural processes to behavior have focused on the processing of motion information in the dorsal stream of the visual cortex [27]. A previous study based on single unit activity in IT failed to establish a link between neural latencies in IT and reaction times variability [28]. The present results, however, directly link neural activity in the ventral stream with real object recognition decisions and their timing. This is in accordance with a more recent single-cell study which demonstrated a correspondence between neural activity in IT and the speed of recognition [29] using isolated real objects in visual search display.

The present study is consistent with theories postulating two distinct modes of visual processing, i.e., an early feedforward processing preserved by the presentation of a backward (pattern) mask, which only interferes with later recurrent processing [8, 12, 30]. However, our analysis further demonstrates that the timing of this initial feedforward sweep is subject to significant modulation consistent with task-specific pre-activation of the ventral stream [31, 32] and/or rapid recurrent feedback within or across areas [33] on slower trials.

Overall, these results suggest that processing based on a rapid feedforward sweep of neural activity through the ventral stream of the visual cortex offers a parsimonious explanation for the very fast behavioral responses observed during rapid categorization tasks.

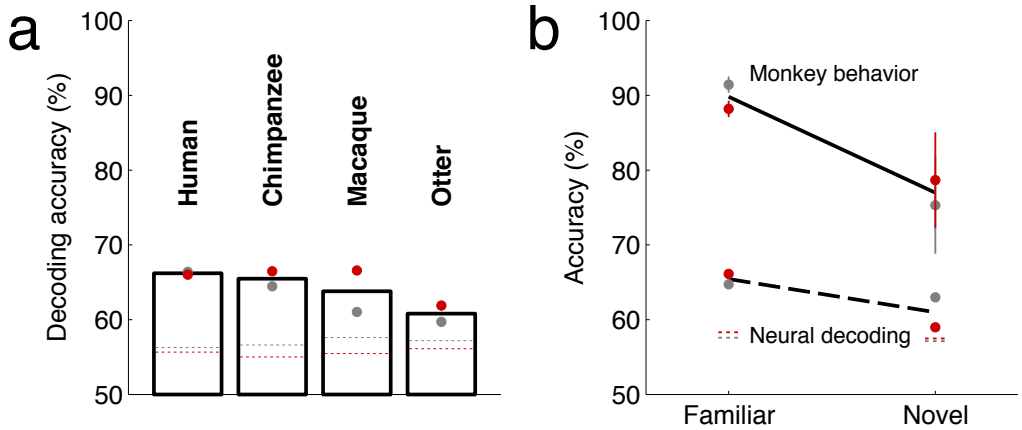

Figure 4. Early ventral stream neural activity: Fine-level visual information and generalization. a) MVPA accuracy for the decoding of animal subcategories (Human, Chimpanzee, Macaque and Otter faces) for target images. Decoding accuracy (90-140 ms time window; 95% confidence interval shown as error bar) based on a one vs. all classification procedure for M1 (gray) and M2 (red). b) Behavioral and neural decoding accuracy on a familiar vs. novel set of stimuli. Data from individual monkeys shown as colored dots with error bars indicating the 95% confidence intervals. Horizontal dotted lines indicate chance level (upper limit of the 95% confidence interval).

Experimental Procedures

Protocol.

Two male rhesus macaques (M1 and M2, both aged 14) performed the experiment. The animals were restrained in a primate chair (Crist Instruments, GA), sitting ~30 cm away from a 1024 x 768 tactile screen. Stimuli were flashed centrally for 33 ms covering about 7° of visual angle on a black background, with a 1.5-3 s random inter-trial interval between successive images. In masked trials (presented in separate blocks), a pattern image corresponding to visual pink noise was presented 50 ms post stimulus onset and lasted 33 ms. Brief presentations, in addition to masking on specific trials, prevented exploratory eye movements and constrained the time available for information uptake.

The two monkeys performed a scene categorization task [5] by releasing a button and touching the screen when they saw an animal in the stimulus presented (target) or keeping their hand on the button otherwise (distractor stimulus). A drop of fruit juice rewarded correct responses; on incorrect trials, the stimulus was re-displayed for 3 s, thus delaying the next reward and motivating the animal to answer as fast and accurately as possible. All procedures conformed to French and European standards concerning the use of experimental animals; protocols were approved by the regional ethical committee.

Stimulus set.

The image set consisted of natural gray-scale photographs (256 256 pixels) equalized for average luminance and global contrast (root mean square over pixel intensities). A "familiar" stimulus set consisting of 280 images (140 animals targets and 140 non-animals distractors) was used to train the animals. In addition, a "novel" stimulus set, never seen before by the monkeys, was used to test for generalization. The novel stimulus set contained 400 stimuli total divided equally between animal target and non-animal distractors. It was introduced to the monkeys in five recording sessions on five consecutive days. On each recording session, monkeys were tested using 1/5 novel stimuli (80 stimuli per day selected at random without replacement from the novel set) and 4/5 familiar stimuli (selected at random from the familiar set). In this novel set, animal targets and non-animal distractors were matched in terms of the distance from the camera [7]. For both stimulus sets, an effort was made for target and distractor scenes to be as varied as possible. Targets included pictures of mammals, birds, insects, reptiles, or amphibians either presented in isolation or in groups and at various positions in the images. Distractors included pictures of various objects (trees, flowers, tools, etc.) in natural and man-made environments (mountain, sea, city, etc.).

Behavioral data analysis.

To test whether monkeys accuracy for the recognition task was above chance, we used a χ^2 test against the null hypothesis that the number of correct and incorrect responses are equiprobable. To compare the animals accuracy on familiar vs. novel and mask vs. no-mask conditions, Pearson's χ^2 test with Yates' continuity correction was used. These analyses were performed for each monkey separately using the R software (available at http://www.r-project.org/).

To compute temporal cumulative d' scores, which measure perceptual discriminability based on behavioral responses, we binned independently hits and false alarms according to reaction times in 30 ms bins and then computed d' scores on these cumulative histograms. Minimal reaction times were computed as the first time bin of at least five consecutive bins for which the number of hits significantly outnumbered the number of false alarms (binomial test; $p<0.01$).

IFP recording and preprocessing.

Monkeys were implanted with subdural macro-electrodes. The implantation location of these electrodes was previously determined using anatomical MRI scans. Under anesthesia, 1 mm holes were drilled in the animal skull and electrodes were lowered manually. The two monkeys where implanted in both hemispheres (9 and 13 electrodes for M1 and M2, respectively). Electrode locations were subsequently assessed by merging X ray (showing electrodes and skull) with T1-weighted MRI anatomical scans (see Fig. S2 for electrode locations). Macroelectrodes were steel wires with 150 microns of diameter and less than 1 MegaOhm impedance. All electrodes were connected to a DB plug. The entire system was fixed on the monkeys skull using screws and dental cement. Recording was performed using NeuroScan EEG system and SynAmps amplifier system sampling at 1000 Hz (band pass 0.1-200 Hz).

Frontal electrodes were used as reference (intra-cortical for M1, within the frontal sinus for M2). M1 was recorded head free, M2 head was restrained. The recording sessions, for each monkey, spanned over 5 consecutive days (3-4 blocks of 600 trials daily until monkey water satiation), resulting in the collection of 7,499 and 8,757 trials for M1 and M2 respectively.

EEGlab toolbox [34] was used to import data in MATLAB and preprocess the signal. 50 Hz noise was removed using a notch filter. Trials used for subsequent analyses were selected based on both potentials (<3 standard deviations from the mean) and global ERP power (sum of square of all electrode potentials for every time point); 10% were discarded based on weak visual power between 60 and 120 ms after stimuli onset. The signal was baseline corrected (50-0 ms) trial by trial.

Computational model.

The Hmax feedforward architecture uses a hierarchy of pooling and template matching layers to account for the increasingly complex and invariant response of neurons along the

ventral stream. The model was shown to match the level of performance of human participants in a rapid natural scene categorization task [7]. Here we used the same parameters as in Serre et al. [7] leading to 7,168 model unit responses for every image (2,048 units for the C1, C2 and C2b layers and 1,024 C3 units). These model unit responses were subsequently fed to a linear classifier as done for the neural data decoding (see below).

Multi-Variate Pattern Analysis (MVPA).
MVPA was performed directly on the neural signal in two ways. A global decoding accuracy measure was obtained by feeding a classifier with the entire 90–140 ms IFP waveforms from all electrodes. A temporal decoding accuracy measure was obtained by feeding each time bin of the IFP signal from all electrodes. Such temporal decoding characterizes the temporal evolution of categorical signal.
For both neural decoding and the computational model, a linear Support Vector Machine (SVM) classifier was used. The classification procedure ran as follow: (1) The image set was equally split in a training and a test set that contained an equal proportion of target and distractor images (2) An optimal cost parameter C was determined through line search optimization using 8-fold cross-validation on the training set of images (3) An SVM classifier was trained and tested on each split. The reported results correspond to the average performance (non-parametric 95% confidence intervals of the mean) using a cross-validation procedure (n=300) whereby different training and test sets were selected each time at random. A measure of chance level was obtained by performing the same analysis on permuted labels. The decoding was considered above chance when 95% of the differences between decoding on true or permuted labels paired over the 300 cross validation was above zero. In the case of temporal decoding, to correct for multiple comparisons, a bin was considered significant when followed by at least five consecutive significant bins [21] ($p < 0.01$).

Animalness scores.
The animalness scores reflect how likely each individual image in the stimulus set is to be classified as an animal irrespective of its actual category label. Here we used this score to compare classification based on neural, behavioral and model data (Fig. S3). An animal score was computed from behavioral data by considering the fraction of trials for which this specific image was classified as (animal) target by the monkeys. For both neural and model data, an animalness score was computed by considering the decision value of the linear (SVM) classifier used for MVPA. The output of the classifier for a specific image corresponds to the distance between this image and the separating hyperplane and can be used as a direct "discriminability measure". Classifier outputs were thus computed for each image of the test set over 50 cross-validations, and then averaged to obtain one single animalness score per image. Animalness scores across all stimuli for behavioral, neural and model data were directly correlated using both Spearman correlation (reported as r^2) and partial correlation (r^2*) measures (using MATLAB built-in functions) to control for category labels. Such partial correlation measures reflect the intrinsic correlation between, for example, the neural and model data, due to visual information beyond predicted category labels. Using this correlation measure, it is thus possible for two visual systems to exhibit a similar level of accuracy and be uncorrelated.

Acknowledgment
All authors contributed to the design of the experiment. M.C. and D.F. trained the monkeys, performed the animal surgery and collected the neural data. M.C., S.M.C. and T.S. conceived and performed the data analysis and wrote the manuscript. All authors approved the final version of the manuscript.
We would like to thank several of our colleagues for their valuable inputs on this manuscript: N. Bichot, D. Brooks, M. Peelen, T. Poggio, D. Sheinberg and R. VanRullen. The monkey electrophysiology was supported by the FRM and DGA. The analysis of the data and comparison to models was supported by DARPA Neovision2 grant to T.S.. Additional support was provided by Brown University, the Center for Computation and Visualization, and the Robert J. and Nancy D. Carney Fund for Scientific Innovation.

References

1. Biederman, I., Rabinowitz, J. C., Glass, A. L., and Stacy, E. W. (1974). On the information extracted from a glance at a scene. J Exp Psychol *103*, 597–600.

2. Potter, M. C., and Faulconer, B. A. (1975). Time to understand pictures and words. Nature *253*, 437–438.

3. Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.

4. Peissig, J. J., and Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? Annu Rev Psychol *58*, 75–96.

5. Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. Neuroreport *9*, 303–308.

6. Bacon-Macé, N. N., Macé, M. J.-M. M., Fabre-Thorpe, M. M., and Thorpe, S. J. S. (2005). The time course of visual processing: backward masking and natural scene categorisation. Vision Research *45*, 1459–1469.

7. Serre, T., Oliva, A., and Poggio, T. A. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U.S.A. *104*, 6424–6429.

8. Vanrullen, R., and Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. J Cogn Neurosci *15*, 209–217.

9. Keysers, C., Xiao, D. K., Földiák, P., Fo, P., and Perrett, D. I. (2001). The speed of sight. J Cogn Neurosci *13*, 90–101.

10. Kovacs, G., Vogels, R., and Orban, G. A. (1995). Cortical correlate of pattern backward masking. Proc. Natl. Acad. Sci. U.S.A. *92*, 5587–5591.

11. Rolls, E. T., and Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proceedings. Biological sciences / The Royal Society *257*, 9–15.

12. Schmidt, T., and Schmidt, F. (2009). Processing of natural images is feedforward : A simple behavioral test. Attention, perception & psychophysics *71*, 594–606.

13. Schall, J. D. (2002). Decision making: neural correlates of response time. Curr. Biol. *12*, R800–1.

14. Johnson, J. S., and Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. Journal of Vision *3*, 499–512.

15. Riesenhuber, M., and Poggio, T. A. (1999). Hierarchical models of object recognition in cortex. Nat Neurosci *2*, 1019–1025.

16. Joyce, C., and Cottrell, G. (2004). Solving the visual expertise mystery. Progress in Neural Processing *15*, 127–136.

17. Fize, D., Cauchoix, M., and Fabre-Thorpe, M. (2011). Humans and monkeys share visual representations. Proc. Natl. Acad. Sci. U.S.A. *108*, 7635–7640.

18. Freedman, D. J., Riesenhuber, M., Poggio, T. A., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science (New York, N.Y.) *291*, 312–316.

19. Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318–320.

20. Dicarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? Neuron *73*, 415–434.

21. Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron *62*, 281–290.

22. Hung, C., Kreiman, G., Poggio, T. A., and Dicarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science *310*, 863–866.

23. Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study.

European Journal of Neuroscience *11*, 1239–1255.

24. Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science (New York, N.Y.) *292*, 510–512.

25. Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R. Q., Adolphs, R., et al. (2011). A category-specific response to animals in the right human amygdala. Nat Neurosci *14*, 1247–1249.

26. Afraz, S.-R. S., Kiani, R. R., and Esteky, H. H. (2006). Microstimulation of inferotemporal cortex influences face categorization. Nature *442*, 692–695.

27. Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

28. DiCarlo, J. J. J., and Maunsell, J. H. R. J. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. J. Neurophysiol. *93*, 2974–2986.

29. Mruczek, R. E. B., and Sheinberg, D. L. (2007). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. J. Neurosci. *27*, 2825–2836.

30. Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. *23*, 571–579.

31. Mirabella, G., Bertini, G., Samengo, I., Kilavik, B. E., Frilli, D., Libera, Della, C., and Chelazzi, L. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. Neuron *54*, 303–318.

32. Peelen, M. V., Fei-Fei, L., and Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature *460*, 94–97.

33. Hupé, J.-M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature *394*, 784–787.

34. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21.

SUPPLEMENTARY FIGURES

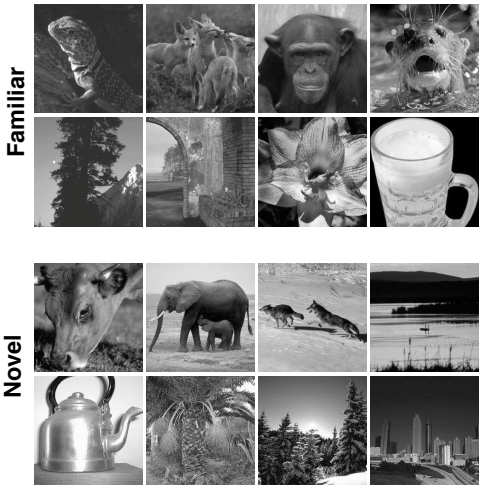

Figure S1. Stimulus sets. Two stimulus sets were used, which consisted in natural gray-scale images with balanced numbers of animal targets and distractors: A "familiar" set consisted in 280 stimuli used to train the animal and a "novel" set consisted in 400 previously unseen stimuli.

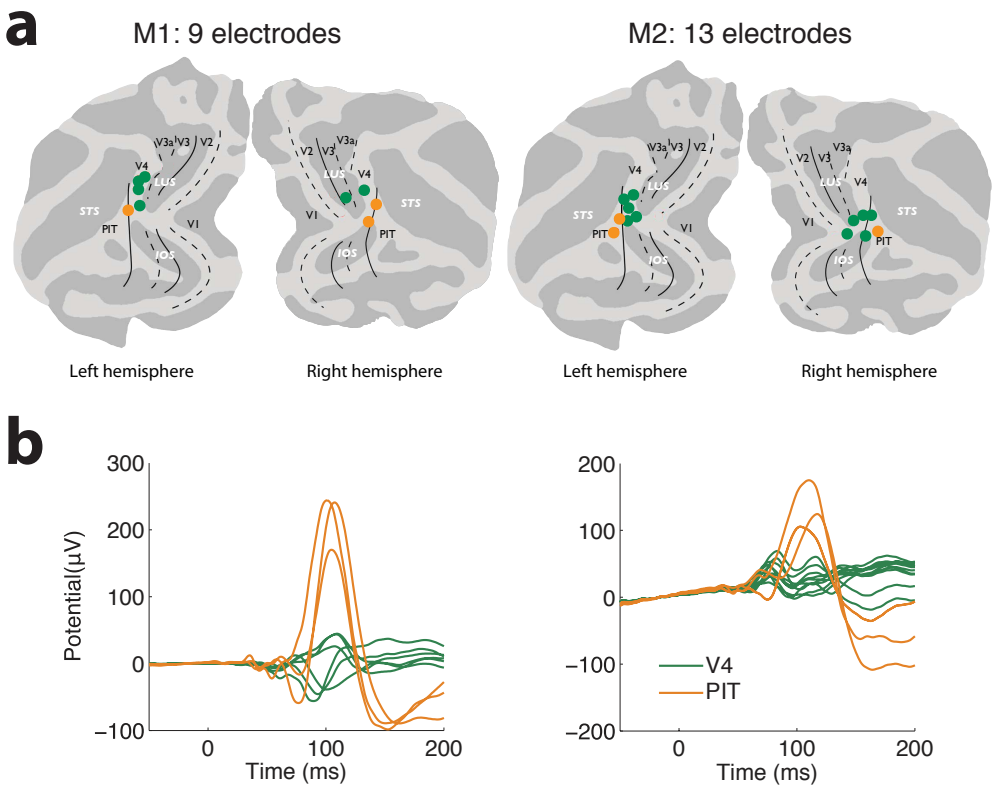

Figure S2. Electrode locations and event related potentials. a) Electrode tip localized on prototypical flat maps obtained from Freesurfer. b) Event Related Potentials (ERPs) averaged over all trials for each electrode (matching colors are used between panel a and b).

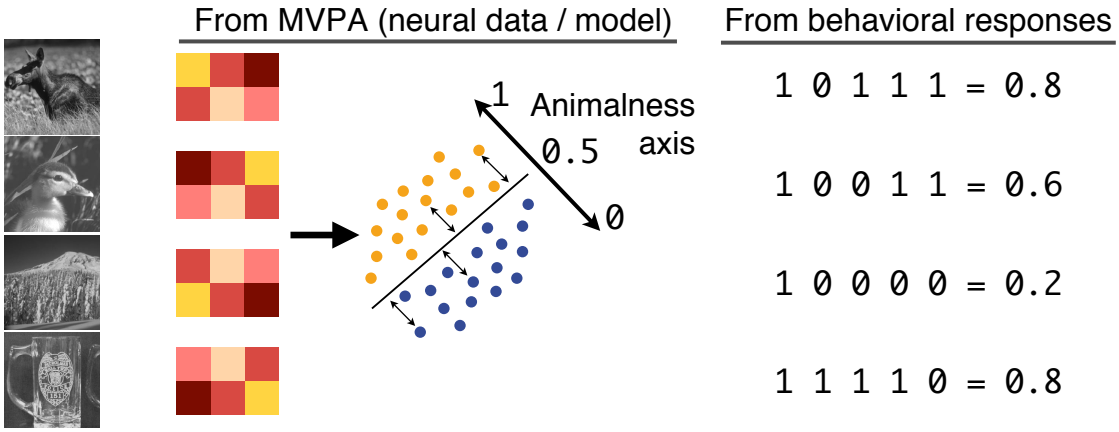

Figure S3. Schematic of the method used to compute animalness scores for individual images. Animalness scores were obtained from neural data and the HMAX computational model by considering classifier outputs / decision values over all cross-validations and averaging them out to get one score per image. For behavioral data, animalness scores were obtained by considering the fraction of go vs. no-go responses over multiple presentations of the same image.

3.4.3 Discussion

La rapidité avec laquelle les primates sont capables de catégoriser des images comme contenant un animal ou non donne un cadre temporel fort pour étudier les bases neuronales de l'abstraction visuelle. Ici, nous montrons que chez le singe, la voie visuelle ventrale présente toutes les caractéristiques suffisantes pour implémenter rapidement un concept aussi abstrait que celui d'animal. La localisation des électrodes, les latences de décodage, la robustesse au masquage et le corrélation avec un modèle hiérarchique ascendant du système visuel semblent indiquer que la voie ventrale construit cette représentation catégorielle à travers une vague ascendante d'activité neuronale dans les aires visuelles de complexité intermédiaire que sont V4 et le cortex inférotemporal postérieur.

Ces résultats qui laissent penser que les aires intermédiaires seraient le support neuronal de la décision dans la tâche de catégorisation rapide sont en accord avec ceux de plusieurs études récentes. D'abord, les neurones V4 et de PIT sont connus pour le codage de formes de complexités intermédiaires (Pasupathy 2006) et de textures (Arcizet et al. 2008), des éléments visuels qui peuvent suffire à diagnostiquer rapidement la présence d'un animal dans l'image. Ensuite, les neurones de V4 peuvent aussi encoder la décision catégorielle comportementale dans une tâche de détection de conjonction de traits visuels simples (Mirrabela et al. 2007). De plus, une lésion du cortex préfrontal, établi comme l'aire nécessaire à la décision dans une catégorisation visuelle (Freedman et al. 2001, Thorpe et Fabre-Thorpe 2001), n'affecte pas la capacité à catégoriser ni à généraliser des singes (Minamimoto et al. 2009). Enfin, les latences de sélectivité dans les aires sous-corticales répondant à des stimuli visuels complexes sont bien trop tardives pour pouvoir influencer les résultats rapportés ici (Mormann et al. 2011).

3.5 Codage neuronal des objets et catégories chez l'homme

3.5.1 Entrée en matière

Nous avons vu dans le chapitre "cognition comparée" que dans une tâche de catégorisation rapide, hommes et singes semblent partager des représentations visuelles très similaires permettant d'accéder au concept d'animal en une fraction de seconde. Le chapitre de méthode exposait quant à lui de nouvelles méthodes d'électrophysiologie et d'analyse du signal permettant de lire essai par essai les représentations visuelles formées dans le cerveau du singe. Enfin, le chapitre précédent démontre que ces méthodes permettent de montrer, chez le singe, un rôle clé des structures cérébrales intermédiaires de la voie visuelle ventrale dans la formation des représentations leur permettant de résoudre la tâche de catégorisation rapide.

Comme nous venons de le faire chez le singe, est-il possible de lire dans un signal électrophysiologique, enregistré chez l'humain, les représentations visuelles formées pendant une tâche de catégorisation rapide? Si oui, est-ce qu'on peut utiliser ces méthodes pour inférer un fonctionnement homologue dans l'implémentation neuronale de concepts visuels chez l'homme et le singe ?

L'anatomie et le fonctionnement du système visuel humain demeurent moins bien caractérisés que chez le singe. Néanmoins, l'avènement de l'imagerie par résonance magnétique fonctionnelle (IRMf) a permis de mettre en évidence des aires spécialisées pour les traitements de catégories d'objets dans la voie visuelle ventrale. En introduction de ce chapitre, nous commencerons par décrire ces principales aires et leur fonction. Dans la deuxième partie de l'introduction, nous montrerons comment des études pionnières en EEG de surface ont permis de circonscrire le temps de traitement cérébral nécessaire pour réaliser une tâche de catégorisation rapide.
Puis, les résultats de ce chapitre seront présentés en deux parties.

1) Les visages humains sont des stimuli particulièrement importants pour nous et fortement représentés dans le cerveau. C'est pourquoi, dans un premier temps nous utiliserons une tâche de catégorisation rapide de visages, pour tenter de transférer à l'EEG de surface enregistré chez l'homme, les méthodes

développées chez le singe pour lire les représentations essai par essai. Nous montrerons que non seulement il est possible de décoder très tôt une représentation de la catégorie visage humain, mais que la précision de décodage de cette représentation dépend des temps de réaction des sujets.

2) Dans le cas de résistance aux traitements médicamenteux, certaines épilepsies sévères doivent être soignées par ablation du foyer épileptique. Lorsque des méthodes non-invasives (EEG, TEP) n'ont pas suffi pour localiser la zone épileptogène, l'implantation de macroélectrodes directement dans le cerveau des patients est réalisée. Les patients gardent généralement ces électrodes durant une à deux semaines, le temps pour les médecins d'effectuer un diagnostique fiable. Pendant cette semaine d'observation, les patients peuvent participer à des tests cognitifs s'ils le souhaitent. Les enregistrements intracrâniens fournissent une formidable opportunité d'étudier le cerveau humain. Effectivement, si les potentiels de champs possèdent une résolution temporelle aussi bonne que l'EEG de scalp, ils présentent en plus un très bon rapport signal sur bruit et enregistrent un signal focal dont on peut déterminer l'origine précisément. En utilisant la même tâche que celle effectuée par les singes, on montre que l'activité intracrânienne chez l'homme permet de lire essai par essai les représentations associées à la tâche de catégorisation animale. Mieux, cette technique nous permet d'esquisser pour la première fois l'existence d'une homologie neurophysiologique entre l'homme et le singe.

3.5.2 Traitement des objets et catégories naturelles dans la voie visuelle ventrale chez l'homme : apport des méthodes hémodynamiques (IRMf, PET)

Des régions cérébrales spécialisées dans le traitement de catégories spécifiques

En 1995, Malach et al. sont parmi les premiers à présenter des objets visuels et des textures, dont la phase a été randomisée, à des sujets placés dans un scanner IRM. En contrastant ces deux conditions, ils révèlent une activation BOLD spécifique des objets dans les aires visuelles extrastriées, au niveau du *cortex occipital latéral*. Ils nomment cette région LOC pour latéral occipital complexe/cortex (Malach et al.

1995). Comme nous l'avons vu dans le chapitre sur le décodage des représentations, Haxby et al. (2001) ont montré qu'un pattern spatial d'activité BOLD, enregistré dans ces régions, permet de prédire avec une grande fiabilité quelle catégorie d'objets a été présentée au sujet parmi diverses catégories visuelles d'objets naturels tels que des visages, des maisons, des chats, des chaises, etc. Ces résultats sont en accord avec l'idée selon laquelle les objets, quels qu'ils soient, seraient représentés par un codage distribué à travers les différentes aires de la voie visuelle ventrale. Néanmoins, un certain nombre d'études semblent montrer qu'il existerait au sein de cette voie des aires spécifiques dédiées au traitement de catégories bien particulières.

Le traitement des visages, de par son rôle essentiel dans les interactions sociales, a toujours beaucoup animé la communauté neuroscientifique. Il semble donc normal que la première aire spécifique découverte soit impliquée dans le traitement des visages humains (Kanwisher et al. 1997). Cette aire se trouve sur la face ventrale du lobe temporal, au niveau du gyrus fusiforme (**Figure 11**), d'où son nom de FFA (Face Fusiform Area). Un an plus tard, la même équipe met en évidence une autre aire spécifique, située à quelques centimètres au niveau du gyrus parahippocampique: la PPA (Parahippocampal Place Area, Epstein et al. 1998) spécialisée elle, dans le traitement des lieux (maisons, paysages, etc.). Depuis, de nombreux travaux ont porté sur la caractérisation et le rôle de ces différentes aires (pour une revue voir Martin, 2007; Kanwisher, 2010; Mahon and Caramazza, 2011; Op de Beek et al. 2008). On sait maintenant que pour les visages par exemple, il n'y a pas qu'une seule aire spécifique mais un réseau de petites régions, chacune possédant un rôle spécifique (Tsao et al. 2008, 2010, 2011). Parmi ces régions, on parle souvent de la plus postérieure appelé OFA (Occipital Face Area, Gautier et al. 2000). On retrouve aussi non loin une aire qui serait spécialisée dans le traitement des corps: EBA (Extrastriate Body Area, Downing et al. 2001 Astafiev et al. 2004, **Figure 11**).

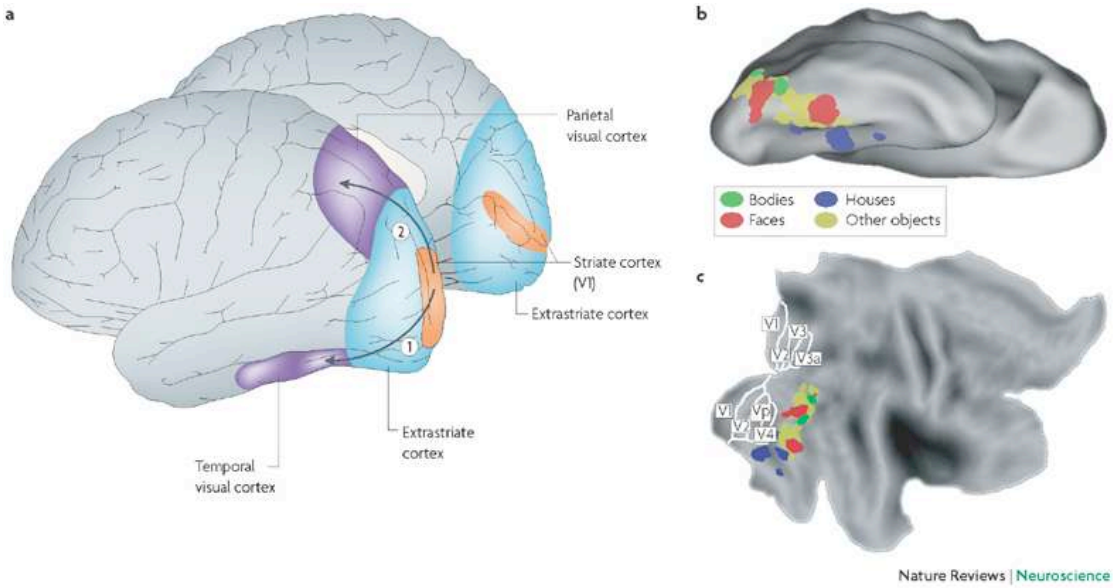

Figure 11: Les grandes subdivisions du cortex visuel humain (A) et les aires de la voie ventrale dédiées au traitement des objets et de certaines catégories d'objets (B, D'après Op de Beeck et al. 2008).

Y a t-il une aire spécifique au traitement des animaux ?

Pour le traitement de la catégorie animale qui nous intéresse particulièrement dans cette thèse, la spécification d'une aire corticale est moins claire. Dans une étude pionnière de tomographie à émission de positron (TEP), Martin et al. (Martin et al. 1996) rapportent que le gyrus occipital médian gauche serait impliqué dans la perception et la dénomination d'animaux (Figure 12).

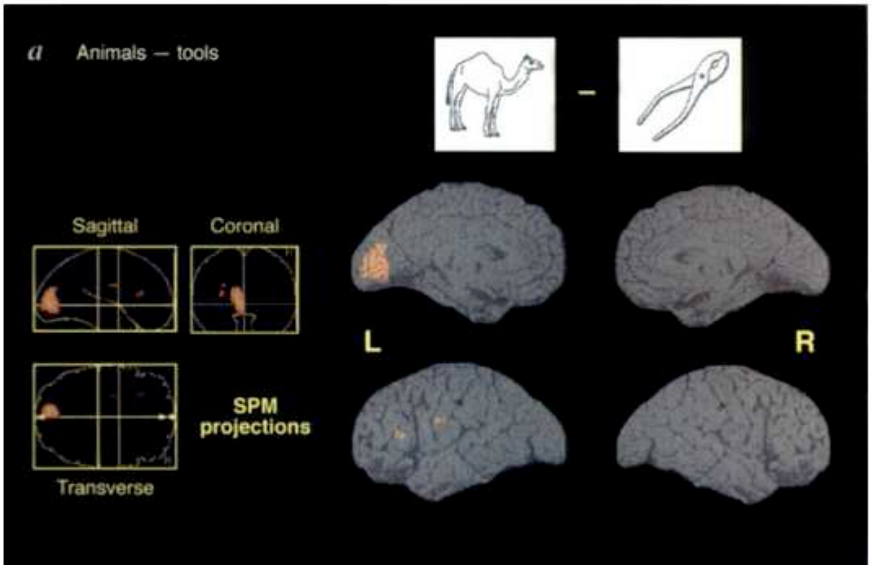

Figure 12 : Un traitement très postérieur de la catégorie « animal » et latéralisé à gauche mis en évidence pour la première fois en TEP par Martin. (Martin et al. 1996)

Une série d'articles confirme un traitement des animaux assez postérieure dans la voie visuelle ventrale impliquant les gyrus médians et inférieurs, dans le lobe occipital et jusqu'au gyrus fusiforme latéral et au gyrus temporal supérieur, dans le lobe temporal (Chao et al. 1999). Wigett et al. (2009) notent l'externalisation des représentations d'objets animés vers la surface du cortex (Figure 13), comparées aux représentations d'objets inanimés, localisées dans des régions plus médiales (gyrus temporal et fusiforme médian).

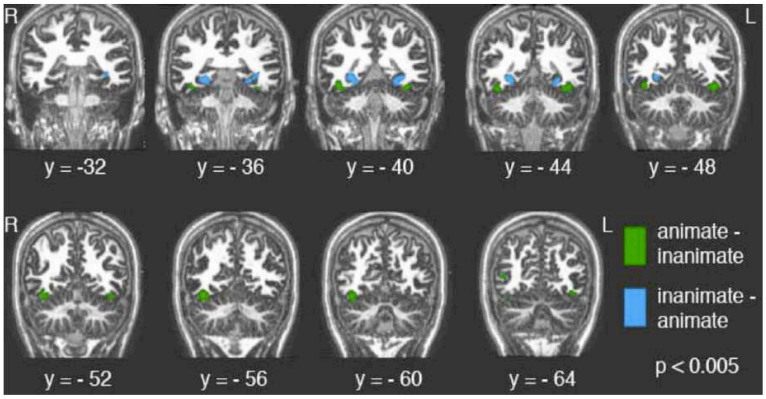

Figure 13: Les images d'animaux activent la voie visuelle ventrale latérale (vert), alors que les objets inanimés activent plutôt la partie médiale (bleue) de manière indépendante à la tâche demandée au sujet. (Wigett et al. 2009)

Plus que dans un simple traitement visuel, ces régions semblent impliquées dans un réel traitement sémantique des animaux et des objets inanimés puisqu'elles s'activent de la même manière dans différentes tâches : d'observation passive, de dénomination ou de mémoire de travail (Wigett et al. 2009). Une étude récente montre que l'activité induite par des sons d'animaux chez des aveugles congénitaux est très similaire à celle induite par la présentation d'images d'animaux chez des sujets contrôles (Mahon et al. 2009). Les auteurs argumentent donc en faveur d'une organisation innée de la sélectivité aux animaux dans ces régions.

En conclusion, il semble que les animaux soient représentés de façon plus distribuée que les visages, corps ou lieux, dans un large réseau très externalisé depuis les gyrus occipitaux médian et inférieur jusqu'au fusiforme. Les animaux étant composés à la fois de corps et de tête, on comprend assez bien pourquoi ils seraient représentés dans un réseau plus large que celui dédié à ces stimuli spécifiques. Effectivement, une étude récente, s'intéressant à la représentation des différentes partie du corps (Orlov et al. 2010), révèle un large réseau occipito-temporal comprenant des sous-régions

impliquées dans le traitement de parties spécifique du corps (torse, bras, jambe, …) qui coïncide largement avec les réseaux mis en évidence dans les études montrant des différences d'activation animé/non animé.

L'activité IRMf semble indiquer un codage précoce (au sein de la hiérarchie visuelle) de la catégorie « animal » qui commencerait dès les premières aires extrastriées. Ces résultats sont en accord avec les réponses rapides observées chez l'humain dans la tâche de catégorisation rapide animal-non animal. Néanmoins, il reste à vérifier que cette activité provient bien d'une vague visuelle ascendante utilisable rapidement pour répondre dans la tâche et non d'un processus descendant plus tardif. Enfin, l'activation de ces régions par des stimuli auditifs chez des aveugles de naissance étaye l'idée selon laquelle ces aires participeraient à un codage du concept multimodal d'animal et ceci de manière pré-câblée. Un tel déterminisme dans le fonctionnement d'un réseau cérébral laisse supposer la possibilité qu'il soit aussi partagé avec d'autres espèces de primates.

Les corrélats de la catégorisation rapide en IRMf.

Une étude récente s'est attelée à révéler les bases neuronales de la catégorisation rapide en utilisant l'IRMf et les méthodes d'analyses multivariées (Peelen et al. 2009). Contrairement à la plupart des études citées précédemment qui utilisent des objets isolés présentés sur fond gris ou même dessinés, dans cette étude des scènes naturelles variées sont présentées aux sujets qui doivent détecter la présence de piétons ou de véhicules malgré l'apparition d'un masque rétrograde. Les résultats pointent le LOC comme substrat probable pour la catégorisation rapide chez l'homme Effectivement, cette région fournit un signal sélectif à la catégorie d'objets présentés dans des scènes naturelles, même pour des zones du champ visuel ne bénéficiant pas d'attention spatiale. Ces résultats sont en accord avec des études comportementales montrant que dans cette tâche, les traitements sont massivement parallèles (Rousselet et al. 2002 et 2004) et sans attention focale (Li et al. 2002, Poncet et al. 2012).

Le second résultat novateur apporté par cette étude est la dépendance entre l'activité catégorielle rapportée dans LOC et la tâche réalisée par le sujet. L'activité sélective n'est présente que lorsque la catégorie est cible dans la tâche. Une explication possible est l'existence de phénomènes de pré-activation. Effectivement, la même

équipe Peelen et Kastner (2010) a ensuite montré que dans une tâche naturelle de recherche visuelle, même en l'absence de stimuli visuels, le LOC à l'inverse de V1 montre une réponse sélective à la catégorie naturelle. Qui plus est, en réalisant une analyse sur le cerveau entier, les auteurs montrent aussi une zone dans le cortex préfrontal médian présentant le même type d'activité. Ces résultats laissent penser que la sélectivité dans les aires visuelles serait liée à des pré-activations induites par le cortex préfrontal. Néanmoins, la résolution temporelle de l'IRMf n'est pas suffisante pour préciser la dynamique entre codage de la tâche et de la catégorie visuelle. Nous espérons clarifier ce dialogue entre aires sélectives et aires impliquées dans le traitement des règles grâce aux enregistrements intracrâniens. Cependant, les résultats de cette étude sont encore trop préliminaires pour être présentés dans ce mémoire de thèse.

3.5.3 Traitement cérébral rapide de catégories naturelles : apports de L'EEG de surface

En 1996, à peu près au moment où Kanwisher, Martin, Haxby et d'autres montraient un codage spécifique des objets dans la voie visuelle ventrale chez l'homme en IRM fonctionnelle, Thorpe et al. publient un papier faisant date sur la vitesse à laquelle les humains et leur cerveau sont capables d'extraire et de traiter une catégorie d'objets à partir de scènes naturelles (Thorpe et al. 1996). En utilisant l'électroencéphalographie sur des sujets engagés dans la tâche de catégorisation rapide, ils montrent qu'on enregistre une activité significativement différente dès 150 ms entre les potentiels évoqués par des images contenant un animal (cible dans la tâche de catégorisation) et celles n'en contenant pas (distracteurs, Figure 14).

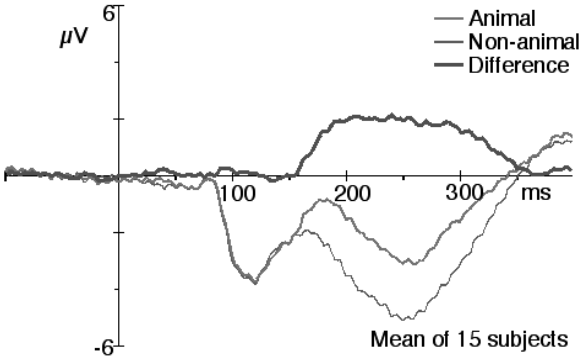

Figure 14: Une activité différentielle entre essais corrects sur les images contenant des animaux cibles et des non-animaux distracteurs apparaît en potentiels évoqués dès 150 ms après le début de la présentation du stimulus (Thorpe et al. 1996)

Ces résultats sont extrêmement importants puisqu'ils fournissent un cadre temporel d'interprétation du fonctionnement du système visuel. Effectivement, si le cerveau peut résoudre une tâche aussi complexe en 150 ms et si les sujets sont capables de répondre en seulement 250 ms, on peut sérieusement questionner la pertinence d'études sur la reconnaissance d'objets rapportant des corrélats neuronaux enregistrés sur 500-1000 ms après présentation du stimulus, comme souvent en enregistrements unitaires, ou même sur plusieurs secondes, comme en IRM fonctionnelle. Ces résultats ont été critiqués et on a en particulier pointé du doigt qu'il pouvait y avoir des différences bas niveau entre les images contenant des animaux et les autres (Olshausen et al. 2003). Mais Vanrullen et al. avaient déjà montré que cette activité différentielle à 150 ms n'était pas liée à la nature visuelle des stimuli mais reflétait la nature cible-distracteur (catégorie abstraite basée sur une règle, VanRullen et Thorpe 2001a). Cette différence n'est pas liée non plus à la couleur puisqu'elle persiste lorsque des stimuli achromatiques sont présentés (Macé et al. 2005). Comme dans l'étude IRMf de Peelen et al. (2009), cette différentielle précoce n'est pas affectée par le traitement parallèle de plusieurs objets (Rousselet et al. 2002).

Rousselet et al. ont aussi étudié les corrélats de la catégorisation rapide de visages humains présentés en contexte dans des scènes naturelles en les comparant à des distracteurs neutres ou à des visages d'animaux (Rousselet et al. 2004). Ils s'intéressent en particulier à la N170, interprétée comme spécifique des visages humains (Bentin et al. 1996, Rossion et Jacques 2008). Or les auteurs rapportent bien une N170 plus forte pour les visages humains par rapport à des stimuli neutres mais qui n'est pas plus ample que celle élicitée par des visages d'animaux. La seule différence significative qu'ils observent est un décalage du pic de la N170 pour les visages animaux comparé aux visages humains. Est-ce que la différence d'amplitude classique de la N170 n'apparaîtrait qu'avec des visages découpés isolés sur fond gris et serait anéantie en présentant des images naturelles ? Ou bien est-ce que l'identification de conspécifiques humains au milieu d'autres visages animaux nécessite un traitement plus élaboré et donc plus lent que celui dont témoigne la N170? Ou finalement, est-ce que la méthode d'analyse des ERPs n'est pas suffisante pour révéler un traitement aussi fin (Rousselet et al., 2007)?

3.5.4 Problématique générale

Comme nous venons de le voir, le traitement des catégories d'objets chez l'humain est traditionnellement abordé à travers deux méthodes : l'IRMf et l'EEG. L'IRMf a permis de mettre en évidence des aires bien spécifiques dédiées au traitement de certains stimuli. Cette technique fournit des patterns d'activation spatiale riches qui, comme nous l'avons vu, permettent grâce à l'analyse de signaux multivariés de lire les représentations mentales, essai par essai. Néanmoins, non seulement ces activations ne reflètent pas directement le traitement neuronal, mais de plus, les latences des modifications de flux sanguin observées sont très en retard par rapport à celles de la perception et catégorisation des objets.

L'EEG permet de révéler des corrélats neuronaux compatibles avec les temps de réponse des sujets, mais semble être un signal grossier ne permettant pas de lire les représentations ni de les localiser. Néanmoins peu de gens se sont « frottés » à l'analyse des signaux multivariés et au décodage des représentions visuelles en EEG dans le domaine de la reconnaissance d'objets. Est-ce que les méthodes que nous avons développées pour décoder les représentations chez le singe pourraient être transférables à l'EEG de surface chez l'homme? De plus, il existe une approche d'électrophysiologie intracrânienne chez l'homme dans le cas d'épilepsie résistante, qui permet de combiner les avantages de l'IRMf et de l'EEG. En effet, elle capture des phénomènes avec une grande précision dans le temps mais aussi dans l'espace. Est-ce qu'une telle méthode peut permettre de révéler un réseau spatio-temporel dans le traitement des catégories et l'émergence de représentations associées?

3.6 Article 4 : Décodage de la catégorie de visage en EEG de surface

Introduction

L'approche classique pour le traitement des signaux électroencéphalographiques est celle des potentiels évoqués : en moyennant l'activité évoquée par de nombreux essais, on peut déduire le temps de traitement nécessaire au cerveau pour distinguer deux types de stimuli (Thorpe et al. 1996).

Néanmoins, l'émergence d'un nouveau champ de recherche en neurosciences pour pallier les handicaps, celui des interfaces cerveau-machine, laisse entrevoir que l'électroencéphalographie pourrait contenir bien plus d'informations que ce que l'on pensait. En effet, en utilisant par exemple le potentiel émis vers 300 ms lorsqu'on est surpris par un événement (Sutton et al. 1965), plusieurs équipes ont montré qu'il était possible, avec des flashs lumineux, de déterminer précisément l'endroit où un sujet porte son attention sur un écran, et ceci en utilisant un très petit nombre d'essais. Fort de ce résultat, en présentant un alphabet sur l'écran, les auteurs ont montré la faisabilité d'un « clavier mental » où le sujet peut écrire des mots en utilisant sa seule activité cérébrale (Farwell et Donchin 1988), un résultat majeur pour les personnes souffrant d'un syndrome d'enfermement leur interdisant tout acte de production motrice ou langagière. Cette idée de suppléance mentale en utilisant l'activité EEG sur un petit nombre d'essais a explosé avec l'augmentation de puissance de calcul des ordinateurs et l'utilisation des techniques d'apprentissage artificiel décrites dans le chapitre sur les méthodes.

A l'instar de ce qui a été réalisé dans le domaine des interfaces cerveau-machine, est-on capable de lire essai par essai les représentions visuelles en utilisant l'EEG? Plus particulièrement, les représentations associées à un traitement haut niveau des objets visuels comme la catégorie des visages? En utilisant la technique d'ERP classique, deux composantes sélectives aux visages sont généralement rapportées : la P100, piquant positivement entre 80 et 130 ms et la N170, piquant entre 130 et 200 ms après la présentation du stimulus. Le rôle respectifs et les deux étapes de traitement dont ces deux composantes témoignent restent sujets à débat (Thierry et al. 2007, Bentin et al. 2007). Est-ce que la lecture des représentations essai par essai peut nous aider à mieux comprendre le traitement temporel des visages ?

Méthodes

Sujet, tâche et stimuli

Pour répondre à cette question, nous avons utilisé l'activité EEG enregistrée grâce à un système neuroscan 32 électrodes sur 28 sujets sains. Les sujets devaient réaliser une tâche de catégorisation de visages humains, dans laquelle on leur demande de lever le doigt d'un bouton le plus vite possible si une photo d'humain leur est présentée (les distracteurs étant des visages d'animaux). Les stimuli sont des photographies en noir et blanc égalisées en luminance moyenne et contraste RMS (270 visages animaux et 270 visages humains).

Prétraitement du signal EEG

On ne garde que les basses fréquences de l'EEG en appliquant un filtre passe bande entre 0.5 et 40 Hz, les hautes fréquences enregistrées sur le scalp ayant souvent été associées à du bruit généré notamment par des mouvements des yeux rapides (Yuval-Greenberg et al. 2008). On utilise la méthode de décomposition en composantes indépendantes (ICA) pour enlever les clignements des yeux. Puis le signal est fenêtré autour de la présentation des stimuli [-100 400] ms et on retranche la moyenne de la ligne de base [- 100 0] ms à chaque pas de temps. Enfin, on sous-échantillonne le signal à 128Hz.

Les méthodes d'analyses multivariées, essai par essai, sont similaires à celles utilisées dans l'article 3.

Principaux résultats

L'analyse essai par essai nous permet de montrer que :

- de l'information relative à la catégorie "visages" est présente très tôt dans le signal (<100 ms après la présentation du stimulus).
- il y a significativement plus d'informations sur les visages pendant la période de la N170 que pendant la période correspondant à la P100.

De plus, en menant le même type d'analyse par quartile de temps de réaction on constate que :

- l'information catégorielle de la N170 est liée au temps de réaction des sujets,
- en revanche il n'y a pas de corrélation entre les temps de réaction et l'information contenue dans la P100.

La même analyse, pas de temps par pas de temps, révèle que cette modulation de l'information relative à la catégorie "visage" liée à la décision du sujet apparaît environ 150 ms après la présentation du stimulus.

ARTICLE 4: Fast EEG decoding of human faces in the wild

Maxime Cauchoix, Gladys Barragan-Jason, Thomas Serre, Emmanuel Barbeau

En préparation

Fast EEG decoding of human face in the wild

Cauchoix M.*[1], Barragan-Jason G.*[1], Serre T.[#2], Barbeau E.J.[#1]
[1] Centre de recherche Cerveau et Cognition, Université de Toulouse, CNRS-UMR 5549, Toulouse, France
[2] Cognitive Linguistic & Psychological Sciences Department, Institute for Brain Sciences, Brown University, Providence, RI, USA
[*,#] equal contributions

The rapid and accurate processing of conspecific face stimuli is critical to primate's social communication. Previous M/EEG studies have suggested that a complex temporal dynamics underlies face processing. In particular, a debate has arisen about the role of an early (P100) vs. late (N170) EEG face selective activity during face processing. To address this issue, we recorded EEG activity while participants performed a rapid go/no-go categorization task in which they had to detect the presence of human faces in complex natural scenes. Single-trial decoding analysis allowed quantifiyng the amount of face category information contained in the EEG signal. Face category information could be readout very early at latencies consistent with P100-related activity. However, we found decoding accuracy from a later N170-related activity to be consistently higher (both within and across subjects). Most importantly, only late N170-related activity but not early P100-related activity was linked to behavioural responses. Altogether these results suggest that, although face category information can be decoded from single trials during an early stimulus-related (P100) EEG activity, behavioural responses are effectively driven by a later decision-related (N170) EEG activity.

Introduction

Primates are inherently social animals that devote much of their visual experience to the processing of faces. Previous electrophysiological studies in human have revealed a complex temporal dynamics underlying face processing with two discrete components: an early component (<120 ms post stimulus onset; see Liu et al., 2002; Halgren et al., 2000) vs. a later component (120-200 ms post stimulus onset; see Bentin et al., 1996; 1996; Rossion et al., 1999; Bentin and Deouell, 2000; Eimer, 2000 andRousselet et al.,and2004; Halgren et al., 2000; Liu et al., 2002). These two components are classically referred to as the P100 and N170 because they peak positively around 100 ms and negatively around 170 ms post stimulus onset respectively.

It has been suggested that the early P100 may reflect the processing of low-level image statistics while the later N170 would more directly support face perception (Rossion and Jacques, 2007; Rossion and Caharel, 2011; Vanrullen and Thorpe, 2001). Others have argued for an opposite view, with a P100 associated with the successful categorization of faces (Liu et al., 2002) and the N170 with low-level variability across stimuli (Thierry et al., 2007). The goal of the present study is to answer this outstanding controversy and understand how these two components relate to the processing of low level visual information vs. high level face perception.

Here we recorded EEG activity while participants were engaged in a rapid go/no-go face categorization task. Participants were asked to discriminate between (target) human faces and (distractor) animal faces (see Figure 1) in a rapid categorization task (Thorpe et al 1996). To precisely quantify the face information carried by the EEG activity in single trials, we performed a multivariate pattern analysis (MVPA) across all electrodes. MVPA has been used in past fMRI studies to understand the neural basis of scene and object processing (Haxby et al., 2001; Spiridon and Kanwisher, 2002; Carlson et al., 2011; Cox and Savoy, 2003; Norman et al., 2006; Reddy and Kanwisher, 2007; Peelen et al., 2009; see Tong and Pratte, 2012 for a recent review) as well as electrophysiological data in both monkey (Hung et al., 2005; Meyers et al., 2008; Zhang et al., 2011) and human (Liu et al., 2009). However, decoding studies from scalp EEG have mainly been limited to simple odd-ball detection tasks in RSVP (rapid serial visual) paradigms (Gerson et al, 2005, Luo et Sajda 2009).

Here we report on the successful decoding of face category information from EEG data from single trials. We found that it was possible to reliably decode the presence or absence of a human face in complex natural scenes very early at latencies consistent with P100-related activity. However, we also found decoding accuracy from the later N170-related activity to be consistently higher (both within and across subjects) and more directly linked to behavioral responses.

Method

Participants

Fifteen females and thirteen males (n=28, mean age: 25.4 years, SD=3.8 years, 25 right-handed) with normal or corrected-to-normal vision signed informed consent to participate in the experiment.

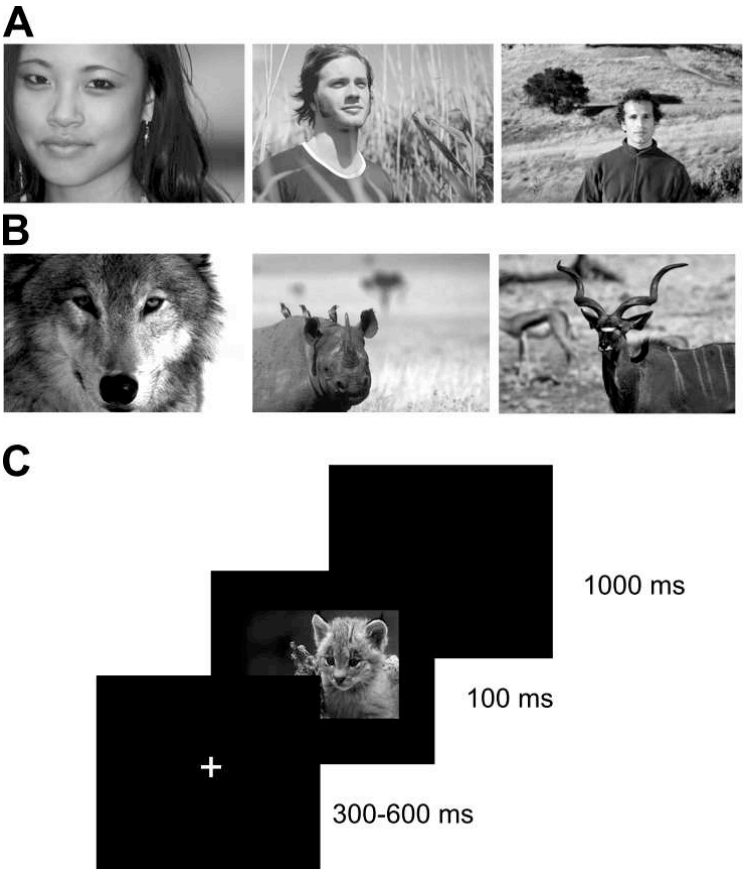

Figure 1: Experimental design.
Examples of target human face (A) and distractor animal face (B) stimuli. (C): One example trial for the go no-go task. After the presentation of a fixation cross for a random time interval (300–600ms), the stimulus was flashed (100ms) then followed by a black screen.

Stimulus set

Gray-scale photographs of (target) human faces (270 images) were presented in their natural context (i.e., images included some background clutter). Animal faces (270 images) including different species (mammals, birds, reptiles, etc; previously used in Rousselet et al., 2003) were used as distractors. Each image was 320 x 480 pixels in size and was controlled for luminance and (RMS) contrast between groups. Examples of photographs are presented in Figure 1.

Experimental Setup

Subjects sat in a dimly lit room at 90 cm from a CRT computer screen. Photographs were displayed on a black background. The experiment consisted in a go-no go paradigm which was divided into three blocks of 180 photographs each (90 targets and 90 distractors). Participants

were trained before performing the task with a specifically dedicated set of stimuli which did not overlap with the stimulus used during the experiment.
Participants were instructed to respond as quickly and accurately as possible by raising their finger from an infrared response pad when a target stimulus was presented (human face / go response). They were asked to keep their finger on the response pad if a distractor stimulus (an animal face) was presented (no-go response). At the beginning of each trial, a fixation cross appeared for a random time interval to prevent as much as possible anticipatory responses (300-600ms). This was followed by the presentation of the stimulus (100 ms) and a black screen (1000 ms, Figure1C). The order of the stimuli presented was randomized across blocks and across subjects.

EEG recording

EEG activity was recorded from 32 electrodes mounted on an elastic cap in accordance with the 10-20 system (Oxford Instruments) with the addition of extra occipital electrodes, using a SynAmps amplifier system (Neuroscan). The ground electrode was placed along the mid-line, in front of Fz, and impedances were kept below 5 kΩ. Signals were digitized at a sampling rate of 1000 Hz and low-pass filtered at 100 Hz. Potentials were referenced on-line to the Cz electrode and average-referenced off-line. EEG data analysis was performed using EEGLAB (Delorme, 2004), a freely available open source toolbox (http://www.sccn.ucsd.edu/eeglab) running under MATLAB (The Mathworks). First, data were down-sampled to 128 Hz and then digitally filtered using a bidirectional linear filter that preserves the phase information (pass-band 0.1–40 Hz). An ICA (Independend Component Analysis) decomposition was used to remove eye blinks for each subject individually. Data epochs were extracted (-100 to 700 ms) and baseline corrected (-100 to 0 ms). A total of 540 epochs for each individual participant was then visually inspected. Epochs containing unique non-stereotypical artifacts (e.g., due to swallowing, electrode cable movements, etc.) were rejected from further analysis. Following this procedure, the mean percentage of rejected epochs was xxx% (mean: xxx, range: xx–xx). For two subjects, one channel was also excluded from further analysis. Only correct trials were used for further EEG analysis, which accounted for xxx% of the overall trials.

Behavioral performance analysis

To estimate the minimal processing time required to detect target images, we computed the shortest reaction time (RT) at which correct go responses started to significantly outnumber incorrect go responses (Rousselet et al., 2003 using an exact Fisher test and 30 ms time bins. To correct for multiple comparisons, the minimal RT was estimated by considering the first significant bin that remained significant for five consecutive bins (60 ms).

ERP analysis

To estimate classical P100 and N170 ERP components, we performed a grand ERP average on T6 (where the P100 and N170 component was maximal) following the method by Rousselet et al. (2003). Statistical difference between human (target) and animal (distractor) trials was assessed using a paired t-test across all 28 subjects for each time point. To correct for multiple comparisons, a difference was reported as significant ($p<0.01$) on five consecutive bins (Liu et al., 2009).

Multi-Variate Pattern Analysis (MVPA)

MVPA was conducted directly on the EEG signal in two ways:

1. A (temporal) decoding analysis was conducted (human vs. animal face) and an accuracy measure was derived by feeding individual time bins of the EEG signal from all electrodes to a linear classifier (see below). Such decoding characterizes the temporal evolution of the category signal.
2. An (epoch) decoding accuracy measure was derived by feeding the entire waveform within a specified time window from all electrodes to the classifier. For this analysis, we considered an early 60–130ms (P100) time window and a later 130–190ms (N170) time window.

For both types of decoding analysis, each feature was normalized (z score) across trials and a linear Support Vector Machine (SVM) was used as classifier (libsvm, Fan et al. 2005). The classification procedure ran as follows: (1) The image set was split equally in a training and a test set that contained an equal proportion of target and distractor images (2) An optimal cost parameter C was determined through line search optimization using 8-fold cross-validation on the training set of images (3) An SVM classifier was trained and tested on each split.

For each subject individually we computed the average performance (non-parametric 95% confidence intervals of the mean) using a cross-validation procedure (n=300) whereby different training and test sets were selected each time at random. A measure of chance level was obtained by performing the same analysis on randomly permuted labels. Decoding was considered significant when readout performance was statistically different from chance (paired t-test across subjects, $p<0.01$). For temporal decoding, as for classical ERP analysis, to correct for multiple comparisons, a bin was considered significant when followed by at least five consecutive significant bins ($p <0.01$).

To get an idea of electrode contribution to the decoding, we plotted the average across subject of the absolute value of raw weights obtained after classifier training.

Results

Behavioral results

Participants performed the categorization task (human vs. animal face) with a very high level of accuracy (mean = 95%, SD = x) and fast reaction times (mean RT = 444 ms, SD = x). The mean minimal RT across subjects was 352 ms . Assuming ~100 ms required for a movement to be triggered (Kalaska and Crammond, 1992, VanRullen and Thorpe, 2001, a decoding analysis based on the first 250 ms following stimulus onset should thus be free of motor artifacts. Thus, even if figures shows data until 400 ms for a comparison with previous ERP studies, we only considered the first 200 ms after stimulus onset for further analysis and statistical testing.

Classical ERP analysis: P100 and N170 components

Visual Evoked Potentials (VEPs) is a classical method to study the time course of face processing. As expected, a large positivity corresponding to the P100 time window (around 110–130 ms, peaking at 115 ms) was observed bilaterally over the lateral occipital sites. This P100 component was followed by a large negativity (around 140–170 ms, peaking at 165 ms) corresponding to the N170 time window over the occipito-temporal sites (Figure 2).

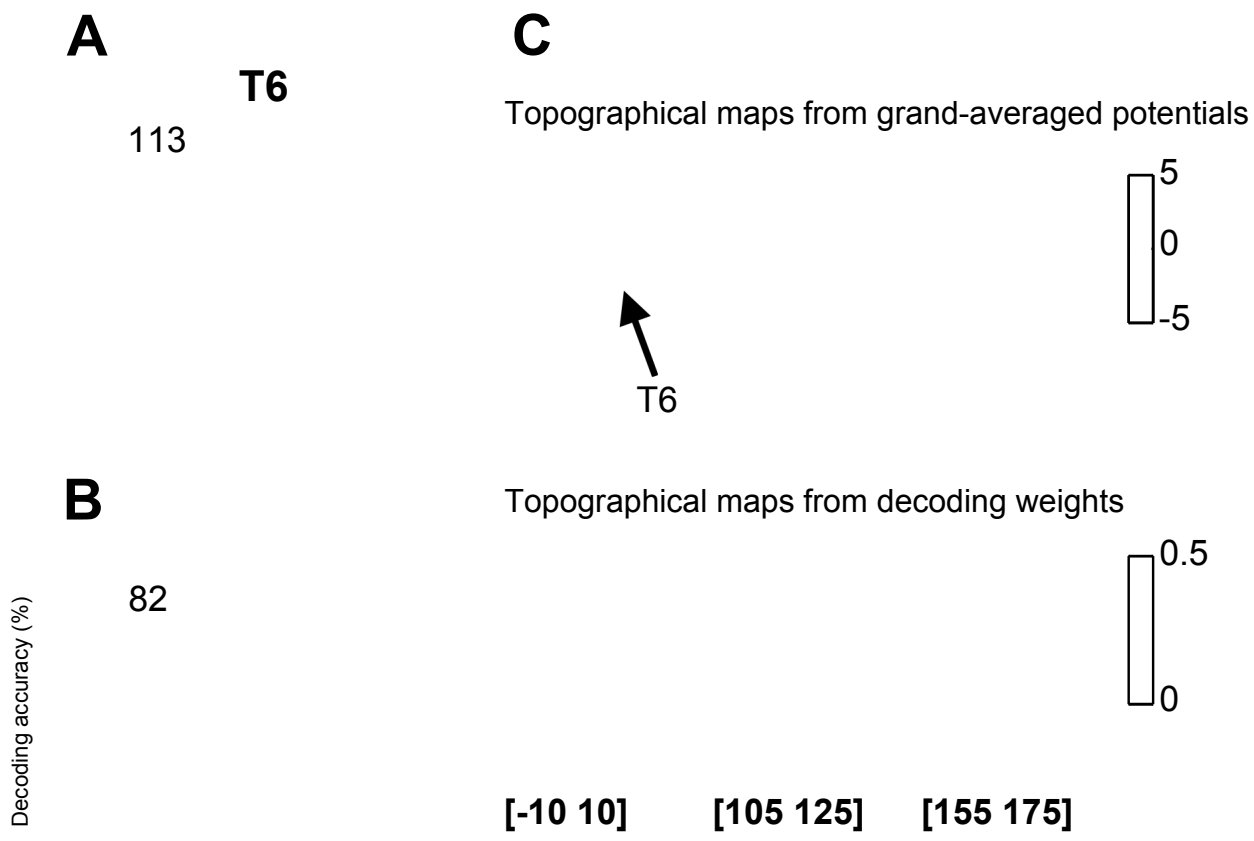

Figure 2: Classical ERP analysis and temporal decoding of face category.
(A) P100 and N170 ERP components. Grand average across all subjects (n=28) on electrode T6 where the N170 component was maximal (see also Rousselet et al., 2004) with target human face trials in green and distractor animal face trials in red. In both conditions, a P100 was observed followed by a N170. The gray bar on the bottom indicates a significant difference between the two conditions. **(B)** Average temporal decoding accuracy across subjects (+/- 95% CI). Vertical dotted line indicates the latency of the first significant decoding above chance level. **(C)** Top: Topographical maps from grand-averaged data across subjects showing the EEG potentials across the scalp (human face trials) for three time windows (-10–10 ms; 105–125 ms; 155–175 ms). Bottom: Normalized MVPA weights of the decoding averaged across participants for the three same time windows.

This negativity was also associated with a positive activity over the central sites. In particular, on T6 (a right occipital-temporal electrode on which both the P100 and N170 amplitudes were maximum), the P100 was significantly higher for animal faces than human faces (paired t-test: $p<0.01$) whereas the N170 amplitude was similar in both conditions ($p>0.05$).

EEG temporal decoding of face category

We further applied MVPA on single trials to better understand how these two components relate to visual processes. Because of the increased sensitivity of multi-variate methods over uni-variate methods, MVPA may better exploit the information contained in the variance of individual trials offering a unique opportunity to link neural processes with behavioral responses. We trained and tested a linear classifier directly on the distributed EEG signal (see Methods). This analysis revealed that face category information is available very early (<100ms post stimulus onset; Figure 2C). Figure 2D shows the resulting normalized weights of the classifier on a topographical map for comparison with the potentials obtained from the univariate ERP

analysis. This analysis suggests that bilateral occipital sites contribute to the decoding occurring during the P100 time window, while decoding during the N170 time window seems to be due to lateralized activity in the right lateral occipital lobe (Figure 2 D).

The N170 carries more face category information than the P100

We further conducted an MVPA separately on the whole 60–130ms (P100) or 130–190 ms (N170) time window (Figure 3). Although decoding was significantly above chance for both time windows (P100: $t(27)=6.5$, $p<0.001$; N170: $t(27)=15.9$, $p<0.001$), decoding was significantly higher for the N170 than the P100 ($t(27)=10.6$, $p<0.001$). As shown on Figure 3, the decoding accuracy was superior for the N170 for each individual subject.

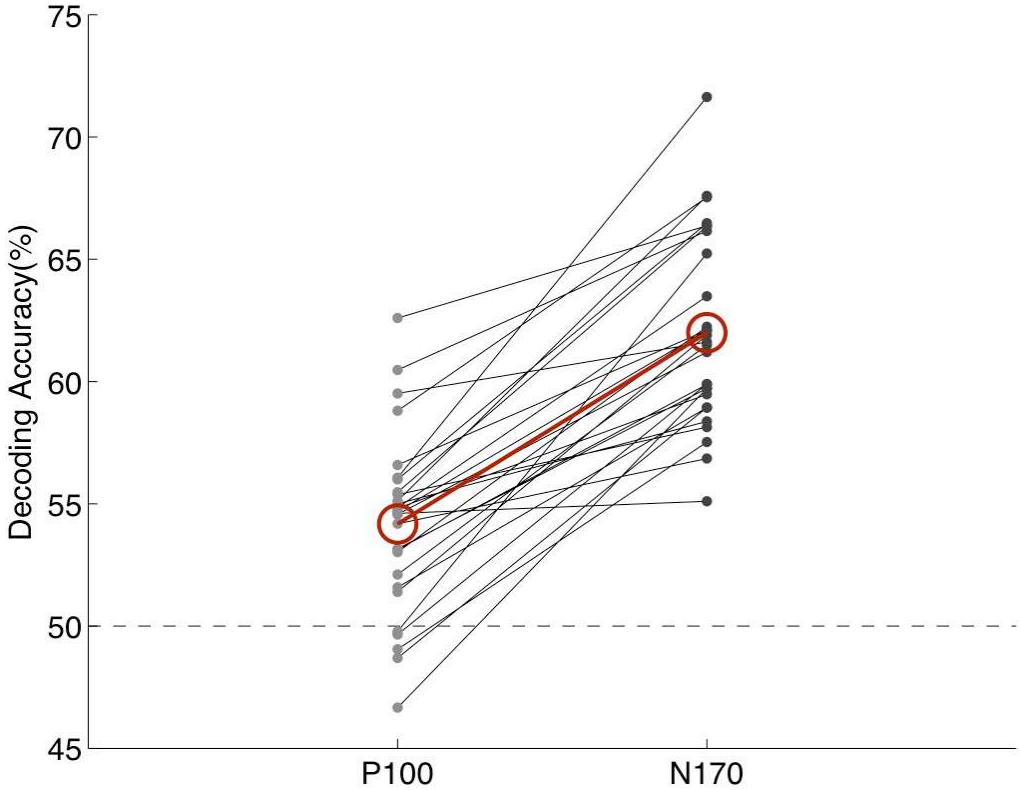

Figure 3: Decoding accuracy during the P100 and N170 time windows.
Decoding accuracy plotted for individual participant (one dot per subject) for the early P100 (gray) vs. late N170 (black) time widow. Mean value for each of time window represented with red circles.

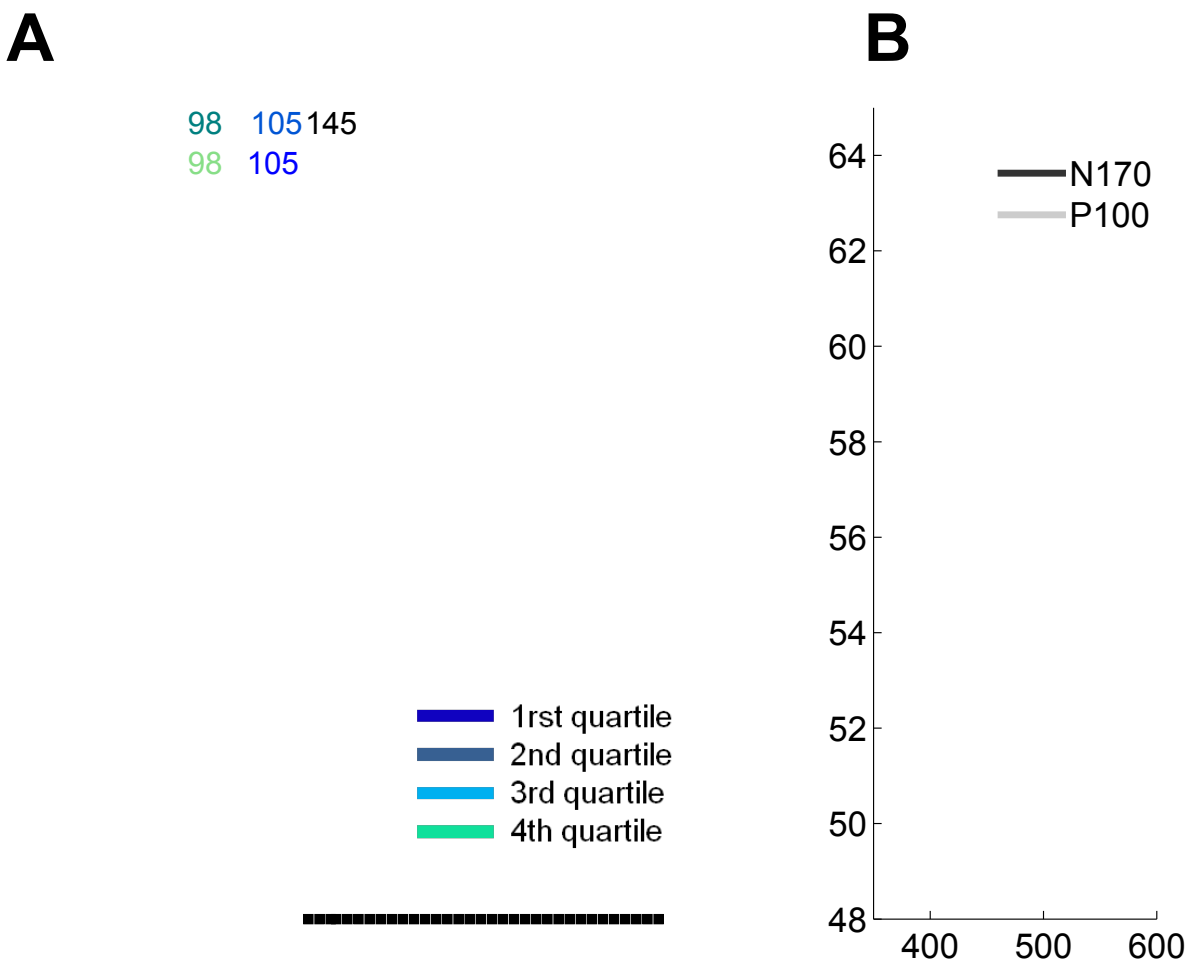

Figure 4: Face decoding and speed of RT. (A) Temporal decoding by quartile. Colored dashed vertical lines indicate decoding onset. Black bar indicates significant difference between the four decoding accuracy. (B) Average decoding accuracy (+/-95% CI) for the P100 (gray) and the N170 (black) time window for individual quartiles of reaction time.

Response time is modulated by N170 but not P100 activity

To test for a link between EEG category activity and response times, we first organized data according to reaction times and computed the decoding accuracy for individual quartiles separately (Figure 4). Temporal decoding for each quartile passed significance at very similar time points (~100ms, Figure 4A). A significant modulation of the decoding accuracy by reaction time only appeared starting at 145 ms after stimuli onset (one way ANOVA, $p<0.01$, 5 consecutive bins). A direct comparison between decoding results for the P100 and N170 time windows suggests a link between decoding accuracy and reaction time for the N170 ($F(3,108)=7.5$, $p=0.001$) but not the P100 ($F(3,108)=1.3$, $p=0.27$) (Figure 4B).

Discussion

What are the neural mechanism responsible for our ability to rapidly categorize faces? M/EEG studies have revealed the existence of two temporal components sensitive to faces: the P100 and the N170. To clarify the debate about the relative role of this two components during face processing, we performed a single trial scalp EEG decoding from human participants engaged in a rapid human vs. animal face categorization task. Our analysis suggests that face category can be read-out as early as 80 ms post stimulus onset. However, it took an additional 65 ms for this category information to start modulating behavioural decisions. How early can object information be read out from single trial EEG? Consistent with a previous study (Vanrullen and Thorpe, 2001), we found early category information < 100 ms that was locked to stimulus presentation. These results seem, however, inconsistent with other MEG studies (e.g. Liu et 2002), which linked P100 activity with face categorisation performance. However, Liu et al. used images in which they varied the phase coherence forcing participants to rely mainly on the power spectrum of these images. it has been shown that the power spectrum of an image specifically modulates the P100 component (Rossion and Caharel 2011) and seems to drive very fast saccades toward faces (Crouzet et al., 2010; Crouzet et Thorpe, 2011). The present study also established a direct link between late face selective information (>140 ms) and behavioural responses. These results seem consistent with earlier decision making studies (Ratcliff et al., 2008; Johnson and Olshausen, 2003). However, the timing revealed by our study was shorter than previous studies (see Rensink et al 2008; Johnson and Olshausen, 2003). These could be explained both by the fact that the statistical power for our MVPA should be higher than for more classical univariate analyis. Furthermore, the rapid categorization paradigm used in the present study may have also constrained participants to operate under more severe time constraints perhaps through a rapid feed-forward sweep of activity (Vanrullen and Thorpe, 2001). Numerous studies have reported a similar category-selective activity between 140-180 ms (e.g. Bentin et al., 1996; Thorpe et al., 1996) leading to the hypothesis that this activity could reflect the earliest high level representation of a stimulus category independently of low level visual information (VanRullen and Thorpe, 2001; Rossion and Caharel; 2011).The results presented here are the first showing that fine categories (human or animal face) of objects presented in complex natural scenes can be reliably decoded from EEG signal using only one single presentation by stimuli. Above all, this single trial information is shown here to directly drive subject behavioural decision in a categorisation task. Surprisingly while MVPA analysis is extensively used for fMRI analysis in cognitive neuroscience, the use of MVPA for M/EEG analysis has been mainly limited to the field of brain computer interface in which people use highly reproducible component such as P300 and use it for various application (P300 speller, Farwell and Donchin 1988). Very few studies have investigated the possibility to read object visual category from non invasive human electrophysiological signal. Among them, a MEG study, while decoding correctly the position of visual objects, shows weak accuracy and late latencies at decoding category of stimuli, even though they used highly different stimuli like isolated faces vs texture or houses (Carlson et al 2011). Another EEG study, reach much better performances to decode line drawing of animals versus tools (Simanova et al 2010). However, the very few number of exemplar used by category, the evident low level visual differences between the two category and the high number of presentation for each stimuli used for decoding let inconclusive about the real nature of decoded information. Altogether, our findings suggest that face processing could rely on a two-step process where low level visual information about the stimulus category is available before 100 ms and information related to decision making is

decoded later but still before 150 ms after the stimulus onset (Thorpe et al., 1996). The first early step could explain low level visual bias affecting the P100 (Rossion and Caharel, 2011) and automatic fast saccades toward faces (Crouzet et al 2011), whereas the second neural step seems to drive behavioral responses recorded during our challenging rapid categorization task related to high level representation of face category.

Reference

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological Studies of Face Perception in Humans. Journal of cognitive neuroscience 8, 551–565.

Bentin, S., Deouell, L.Y., 2000. Structural encoding and identification in face processing: erp evidence for separate mechanisms. Cogn Neuropsychol 17, 35–55.

Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. J Vis 11.

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, 261–270.

Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: face detection in just 100 ms. J Vis 10, 16.1–17.

Crouzet, S.M., Thorpe, S.J., 2011. Low-level cues and ultra-fast face detection. Front Psychol 2, 342.

Eimer, M., 2000. The face-specific N170 component reflects late stages in the structural encoding of faces. Neuroreport 11, 2319–2324.

Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput Assist Interv 8, 1–8.

Farwell, L.A., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 70, 510–523.

Gerson, A.D., Parra, L.C., Sajda, P., 2005. Cortical origins of response time variability during rapid discrimination of visual objects. Neuroimage 28, 342–353.

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., Hari, R., 2000. Cognitive response profile of the human fusiform face area as determined by MEG. Cereb. Cortex 10, 69–81.

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, 2425–2430.

Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J., 2005. Fast readout of object identity from macaque inferior temporal cortex. Science 310, 863–866.

Johnson, J.S., Olshausen, B.A., 2003. Timecourse of neural signatures of object recognition. J Vis 3, 499–512.

Kalaska, J.F., Crammond, D.J., 1992. Cerebral cortical mechanisms of reaching movements. Science 255, 1517–1523.

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, 281–290.

Liu, J., Harris, A., Kanwisher, N., 2002. Stages of processing in face perception: an MEG study. Nat. Neurosci. 5, 910–916.

Luo, A., Sajda, P., 2009. Comparing neural correlates of visual target detection in serial visual presentations having different temporal correlations. Front Hum Neurosci 3, 5.
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. J. Neurophysiol. 100, 1407–1419.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. (Regul. Ed.) 10, 424–430.
Peelen, M.V., Glaser, B., Vuilleumier, P., Eliez, S., 2009. Differential development of selectivity for faces and bodies in the fusiform gyrus. Dev Sci 12, F16–25.
Ratcliff, R., Philiastides, M.G., Sajda, P., 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc. Natl. Acad. Sci. U.S.A. 106, 6539–6544.
Reddy, L., Kanwisher, N., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr. Biol. 17, 2067–2072.
Rossion, B., Caharel, S., 2011. ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. Vision Res. 51, 1297–1311.
Rossion, B., Delvenne, J.F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., Guérit, J.M., 1999. Spatio-temporal localization of the face inversion effect: an event-related potentials study. Biol Psychol 50, 173–189.
Rossion, B., Jacques, C., 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. Neuroimage 39, 1959–1979.
Rousselet, G.A., Macé, M.J.-M., Fabre-Thorpe, M., 2003. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. J Vis 3, 440–455.
Rousselet, G.A., Macé, M.J.-M., Fabre-Thorpe, M., 2004. Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? J Vis 4, 13–21.
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, e14465.
Spiridon, M., Kanwisher, N., 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 35, 1157–1165.
Thierry, G., Martin, C.D., Downing, P., Pegna, A.J., 2007. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nat. Neurosci. 10, 505–511.
Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.
Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. Annu Rev Psychol 63, 483–509.
VanRullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. J Cogn Neurosci 13, 454–461.
Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., Desimone, R., 2011. Object decoding with attention in inferior temporal cortex. Proc. Natl. Acad. Sci. U.S.A. 108, 8850–8855.

Discussion

En utilisant une technique multivariée d'analyse des potentiels de champs, développée chez le singe, on décode ici pour la première fois en EEG de surface chez l'homme des représentations visuelles haut niveau liées à la perception de visages dans des scènes naturelles. Cette méthode d'analyse nous permet notamment d'éclairer le débat sur les rôles des deux composantes rapportées classiquement en ERP. Effectivement, si de l'information relative à la catégorie de visages est déjà présente dans le signal avant 100 ms, celle-ci n'est pas liée à la réponse comportementale des sujets dans la tâche. On peut néanmoins imaginer que cette information disponible très tôt pourrait être utilisée pour réaliser les saccades ultra-rapides et automatiques vers les visages humains récemment rapportés (Crouzet et al. 2010). En revanche, un peu plus tard, autour de la fameuse N170, l'information catégorielle de visage est quantitativement supérieure et corrélée au temps de réaction des sujets, ce qui semble indiquer l'émergence, à ce moment, d'une représentation haut-niveau des visages utilisée par les sujets pour prendre leur décision catégorielle.

Ces résultats sont en accord avec une autre étude EEG dans une tâche de catégorisation animal-véhicule (VanRullen & Thorpe 2001). Les auteurs dissocient l'information visuelle bas-niveau de l'information catégorielle et décisionnelle haut-niveau en contrastant les ERPs enregistrés pendant la présentation des mêmes stimuli, mais lorsqu'ils sont cibles ou distracteurs dans la tâche (différentielle secondaire). Alors que la différentielle classique (animaux cibles contre voiture distracteurs) révèle une différence statistique dès 70-80 ms, la différentielle secondaire (animaux cibles contre animaux distracteurs) n'est significative que 150 ms après la présentation du stimulus.

L'analyse multivariée essai par essai permet non seulement de quantifier précisément la quantité d'informations catégorielles mais elle permet aussi de montrer des différences qui ne sont pas forcément révélées par la méthode classique des ERPs. En effet, dans une étude similaire Rousselet et al. (2004) n'avaient pas montré de différence significative d'amplitude entre potentiels évoqués par des visages humains et des visages animaux en utilisant les méthodes classiques d'analyse EEG. Ils n'avaient pas non plus trouvé d'activité liée à la tâche comme Vanrullen et Thorpe

(2001). Les auteurs proposent que, vu leur importance biologique, les visages seraient traités par défaut jusqu'à un plus haut niveau de détail.

Sans prendre en compte le caractère particulier des scènes naturelles et la différence visages humains, visages animaux, peu d'études rapportent un bon décodage de la catégorie en EEG. Une étude récente en MEG montre de bonnes performances pour décoder la position rétinienne d'objets visuels. Cependant, malgré les 160 canaux d'enregistrement et l'utilisation de stimuli connus pour éliciter des réponses catégorielles très différentes en IRMf (visages et maisons), cette étude observe des pourcentages de décodage faibles (<60%) et tardifs (>130 ms), lorsqu'il s'agit de décoder la catégorie des stimuli (Carlson et al. 2011).

A notre connaissance une seule autre étude a réussi à décoder la catégorie visuelle en EEG (Simanova et al. 2010). Mais cette étude utilisant seulement 4 stimuli par catégorie, il est difficile de parler de catégories avec si peu d'exemplaires. Qui plus est, ces stimuli sont des animaux ou des outils dessinés à la main qui diffèrent beaucoup d'un point de vue bas-niveau visuel : les animaux ont une forme globale tassée (lion, vache, ours, singe), alors que les objets sont allongés (hache, ciseaux, stylo, peigne). Il est fort probable que les très hautes et très rapides performances de décodage qu'ils obtiennent soient dues à de telles différences bas-niveau. Dans notre cas, l'utilisation d'une large banque de stimuli et l'égalisation des stimuli en luminance et contraste rendent moins probable un biais visuel bas-niveau.

Une des possibilités qui pourrait expliquer nos bonnes performances dans la lecture des représentations concerne la tâche "active" que nous utilisons. Nos sujets sont engagés dans une tâche active de catégorisation rapide alors que les autres groupes utilisent soit des tâches de fixation passive, soit des tâches orthogonales. En effet, l'entraînement et l'implication dans une tâche de catégorisation semblent augmenter la sélectivité des aires visuelles aux catégories (Jiang et al. 2007). De plus, la contrainte temporelle mise sur les sujets en flashant les stimuli pendant 100 ms et en leur demandant de répondre en moins de 1s permet peut-être de forcer l'apparition de représentations catégorielles fortes et robustes, plus rapidement et facilement décodables essai par essai.

3.7 A la recherche d'une homologie neurophysiologique fonctionnelle de la catégorisation rapide

Introduction et problématique

Nous avons vu dans le chapitre précédent que, comme chez le singe, il est possible de décoder à partir du signal EEG les représentations précoces que l'homme utilise pour résoudre la tâche de catégorisation rapide. Ces résultats laissent supposer un fonctionnement neurophysiologique sous-jacent commun (au moins partiellement) à l'homme et au macaque.

Néanmoins, la comparaison des deux études (Article 3 et 4) est loin d'être parfaite. D'abord, les stimuli utilisés dans les deux expériences sont différents : alors que les singes font une tâche de catégorisation d'animaux parmi des non-animaux, les hommes font une tâche de catégorisation de visages humains face à des visages animaux. Or, nous avons vu que chez l'homme, les zones corticales spécialisées dans le traitement des visages humains (Kanwisher et al. 1997) diffèrent sensiblement des aires impliquées dans le traitement de la catégorie animale (Chao et al. 1999).

Ensuite, les méthodes d'enregistrement sont différentes. Si chez le singe on enregistre des potentiels de champs intracrâniens provenant d'une zone bien ciblée qui est celle des aires visuelles intermédiaires ventrales, le signal EEG de surface est un signal provenant de l'ensemble du cerveau et dont la diffusion, après la traversée des méninges et du crâne, permet difficilement d'en localiser les sources. Cette différence permet sans doute d'expliquer que, malgré des latences de décodage très similaires chez l'homme et le singe (70-90 ms après présentation du stimulus), on observe un lien avec les réponses comportementales dès le début de la réponse catégorielle chez le singe, alors que cette corrélation n'apparaît que dans une deuxième phase (50-70 ms plus tard) en EEG chez l'homme. Comme en EEG on capte du signal provenant de l'ensemble du fonctionnement cérébral, il est possible que l'on décode dans un premier temps une activité visuelle bas-niveau, non pertinente pour la tâche, provenant des aires primaires (V1/V2) et ce n'est que plus tard que l'on décode une activité catégorielle sous-tendant la décision des sujets et provenant des aires visuelles extrastriées (LOC), impliquées dans le traitement des objets.

Pour passer outre ces différences méthodologiques, nous allons présenter ici des résultats préliminaires obtenus chez le patient épileptique. Dans des cas d'épilepsies résistantes aux traitements médicamenteux, la seule thérapeutique consiste à enlever la partie du cerveau provoquant les crises. Pour cela les médecins épileptologues ont généralement recours à l'EEG de surface qui leur permet, après visionnage du tracé EEG pendant la crise, de localiser le foyer épileptique avant la chirurgie. Dans certains cas, l'EEG de surface n'est pas suffisante et il est nécessaire de recourir à des techniques invasives d'enregistrements intracrâniens, directement à la surface du cortex (électrocorticographie, EcoG) ou en profondeur (EEG stéréotaxique, sEEG) qui permettent une localisation plus précise du foyer. Nous avons eu la chance de faire passer un protocole de catégorisation rapide animal/ non animal, en employant exactement les même stimuli que ceux utilisés chez le singe, à trois de ces patients présentant une implantation ECoG principalement occipito-temporale.

Ces enregistrements chez le patient épileptique peuvent-ils nous permettre de mettre en évidence un réseau cérébral impliqué dans la formation de concepts visuels qui soit homologues aux aires visuelles intermédiaires du macaque Rhésus ?

Hypothèses et prédictions

Dans l'article 1, nous avons insisté sur les ressemblances comportementales entre le macaque et l'homme dans la réalisation de la tâche de catégorisation rapide animal /non animal. Ces similarités justifient la recherche d'une homologie neuronale menée dans ce chapitre. Néanmoins, il existe une différence notable entre les réponses des hommes et des singes : la vitesse de réponse. Effectivement, à en juger par les latences des premières réponses significatives (200 ms pour le singe, 250 ms pour l'homme), il existe un décalage de 50 ms entre les temps de réaction minimaux de l'homme et du singe. Un décalage que l'on retrouve aussi lorsque les sujets réalisent la tâche en saccade, dans laquelle les hommes atteignent un temps de réaction minimal de 150 ms (Kirchner et Thorpe 2005) contre 100 ms pour les singes (Girard et al. 2008). Ce coût temporel entre les deux espèces ne peut donc pas être attribué à la modalité de réponse mais bien aux étapes sensorielle, catégorielle et/ou décisionnelle.

Par conséquent, un réseau neuronal homologue entre le singe et l'homme, impliqué dans le codage du concept d'animal et dont l'activité soutiendrait directement les réponses observées dans la tâche de catégorisation rapide, devraient respecter les critères suivants :

1- Un classifieur devrait pouvoir décoder la catégorie animale, à partir du signal provenant d'une telle structure, 50 ms plus tardivement chez l'homme que chez le singe.
2- Cette activité catégorielle doit être liée aux réponses comportementales de chaque espèce et le début de cette corrélation doit aussi présenter un décalage d'environ 50 ms.
3- Chez le singe, on estime à environ 100 ms le temps nécessaire pour déclencher une réponse manuelle à partir du signal cortical (Kalaska et Cramon 1992) et 20 ms pour une réponse en saccade (Crouzet et al. 2010). Par conséquent, la réponse catégorielle devrait débuter vers 80/100 ms après présentation du stimulus chez le singe (100 - 20 pour les saccades/ 200 - 100 pour les réponses manuelles) et donc 50 ms plus tard chez l'homme soit 130/150 ms après présentation du stimulus.

Méthodes

Les patients on été enregistrés par Thomas Serre et Jed Singer, sous la supervision de Gabriel Kreiman, à l'hôpital pour enfants de Boston aux Etats-Unis. Chaque patient est implanté sous la dure-mère avec des grilles d'électrodes (EcoG). La localisation des électrodes est obtenue en fusionnant l'IRM anatomique préopératoire et un scanner X post opératoire grâce au logiciel libre Freesurfer. Les patients passent la même tâche que celle utilisée pour l'article 3. Dans la moitié des cas les essais sont masqués par un pattern en 1/f. La seule différence avec la tâche passée par les singes, est que les patients répondent avec deux touches (une pour animal, l'autre pour non animal) alors que les singes répondaient en go/no-go (go sur animal). Les procédures d'analyse du signal, de décodage et les statistiques employées ici sont les mêmes que celles utilisées dans l'Article 3. Pour la partie comportementale, afin de comparer les performances des singes avec un nombre suffisamment important de sujets humains, en plus des performances des 3 patients épileptiques, nous rapportons celles de 22 sujets sains enregistrés par Thomas Serre au MIT (Serre et al. 2007).

Résultats

Avant toute analyse du signal neuronal, peut-on retrouver une similarité comportementale entre les réponses des humains et celles des singes dans cette tâche et avec ces stimuli?

Les deux singes atteignent dès la première présentation des stimuli une précision de catégorisation comparable à celle des sujets contrôles et des 3 patients épileptiques, que les stimuli soit masqués ou non (Figure 15 A).

On utilise la mesure *d'animalité* (% de réponse go par image sur les 5 premières présentations pour chaque singe ; et % de réponses oui par image à travers 22 sujets humains contrôle sur 1 seule présentation) pour évaluer la similarité des réponses des singes et des hommes à l'échelle de l'image. La Figure 15 C illustre la corrélation entre l'animalité moyenne des deux espèces. L'évaluation statistique de cette corrélation est réalisée de la manière suivante. Afin de tenir compte de la part de corrélation due à la bi-modalité des réponses (animal ou non animal), on effectue la même corrélation sur 3000 permutations de labels intra-catégorie. La corrélation réelle moins la moyenne de ces 3000 corrélations obtenue après permutation nous donne la valeur de corrélation corrigée, reportée dans la Figure 15 B. La limite supérieure de l'intervalle à 95% de ces corrélations permutées donne le seuil de significativité ($p=0.05$). On observe donc une corrélation significative des animalités inter-espèces. Les valeurs de corrélations d'animalité inter-espèces avoisinent d'ailleurs celles intra-espèces. Les singes et les sujets humains ont donc tendance à considérer les mêmes images comme étant de bons ou mauvais exemplaires de la catégorie animale (voir aussi Fabre-Thorpe et al. 1998).

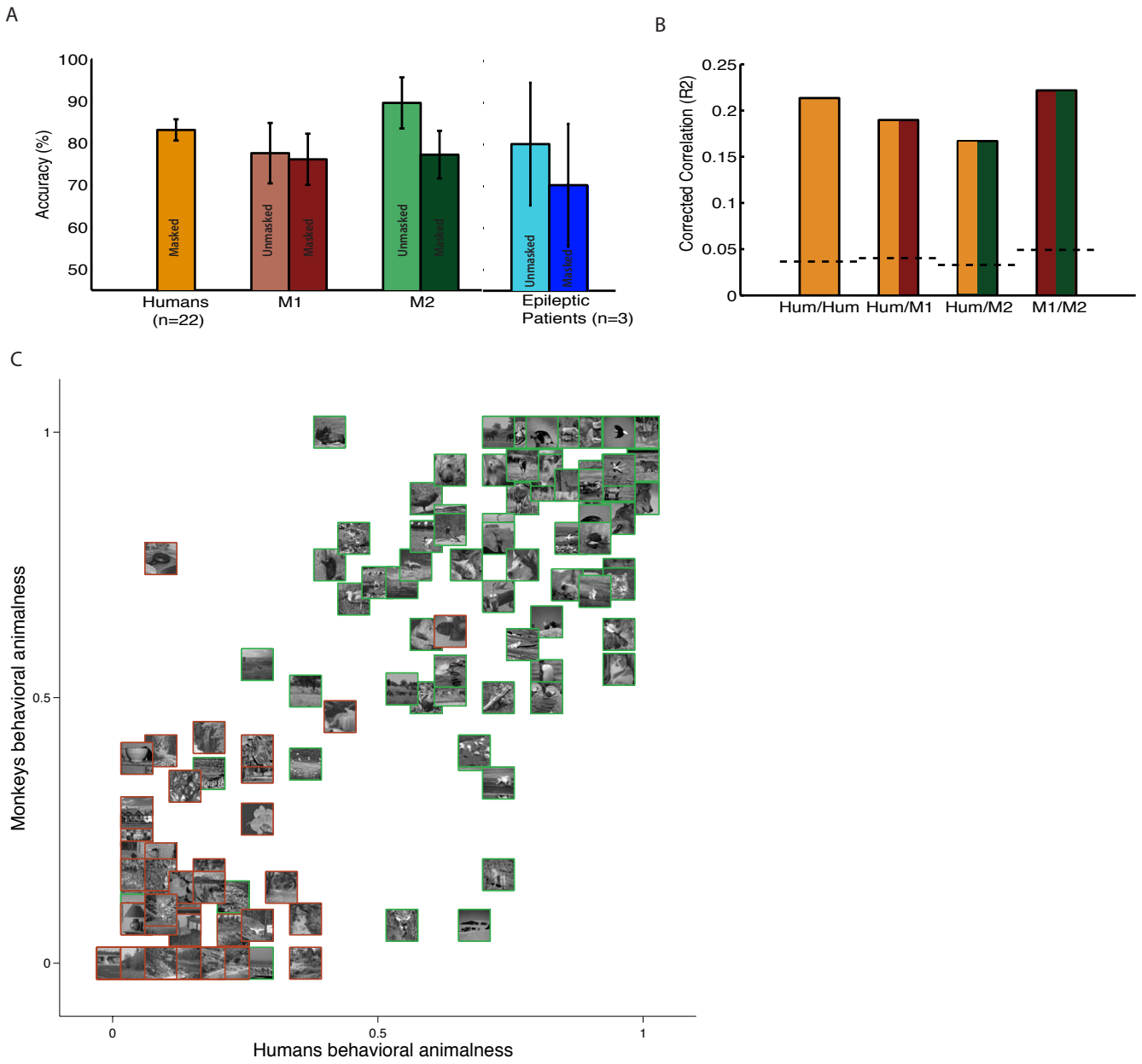

Figure 15: Performances comportementales des 2 singes (M1, M2), des 3 patients épileptiques et de 22 sujets humains contrôles. (A) Précision de catégorisation (+/- intervalle de confiance à 95%) avec ou sans la présence d'un masque rétrograde. (B) Corrélation corrigée (R2 réel – R2 labels permutés intra-catégorie), à travers les images, entre l'animalité comportementale de chaque singe (% de réponse go par image sur 5 premières présentations) et celle des 22 humains contrôles (% de réponses oui par image à travers les 22 sujets sur 1 seule présentation). Le trait horizontal pointillé marque la limite supérieure de l'intervalle à 95% autour de la chance pour cette mesure. (C) Animalité comportementale moyennée entre les deux singes (% de réponse go par image à travers tous les essais des deux singes) en fonction de celle des 22 humains contrôles pour chaque image (cible animale en vert et distracteurs non animaux en rouge).

Comme dans l'article 1, singes et hommes semblent recourir à des représentations similaires pour effectuer cette tâche de catégorisation animal/non animal. En utilisant les outils permettant de lire ces représentations catégorielles dans les potentiels de champs, peut-on mettre en évidence chez l'homme, grâce aux enregistrements acquis chez les trois patients épileptiques, des territoires corticaux homologues aux aires visuelles intermédiaires chez le singe ?

Après juxtaposition du Scanner X sur l'IRM anatomique, on récupère les coordonnées Talairach de chacune des électrodes. Grâce à ces coordonnées, on peut savoir au-dessus de quelle zone corticale était placée l'électrode et donc, en assumant que l'électrode enregistre principalement le territoire sur lequel elle est disposée, la provenance du signal enregistré.

On divise les électrodes de chaque patient en trois régions d'intérêt:

- une région correspondant aux *aires occipoto-latérales sélectives pour les animaux* (qu'on appellera LOAnA pour Lateral Occipital Animal Areas), comme définie fonctionnellement par Chao et al. (1999). Cette région comprend les électrodes situés sur les *gyrus occipitaux inférieurs*, *médians* et les *gyrus fusiformes*.
- une région *temporale* regroupant les *gyrus temporaux inférieurs*, *médians* et les *gyrus parahippocampiques*
- et une *région fronto-pariétale* regroupant toutes les électrodes situés au-dessus des *lobes frontaux* ou *pariétaux*.

Table 4: Nombre d'électrodes par patient pour chaque région d'intérêt.

	LOAnA	**Temporal**	**Fronto-parietal**
Patient 1	11	22	5
Patient 2	16	39	5
Patient 3	16	28	3

On rapporte les performances de décodage, moyennées à travers les sujets, sur les mêmes images exactement, pour les deux singes et chaque région d'intérêt des trois patients (Figure 16 A). Chez l'homme, pendant les 400 ms après présentation du stimulus, on ne décode pas la catégorie avec les électrodes fronto-pariétales. Le

décodage avec le signal temporal, malgré un échantillonnage dense (Table 4), ne devient significatif que 240 ms après la présentation du stimulus. Par contre, le signal provenant de la LOAnA permet de décoder la catégorie très tôt (130ms), seulement 50 ms après la latence du décodage chez le singe (80 ms). En terme de puissance de décodage maximal, la LOAnA et les aires intermédiaires du singe apparaissent très similaires, les deux avoisinant les 62%.

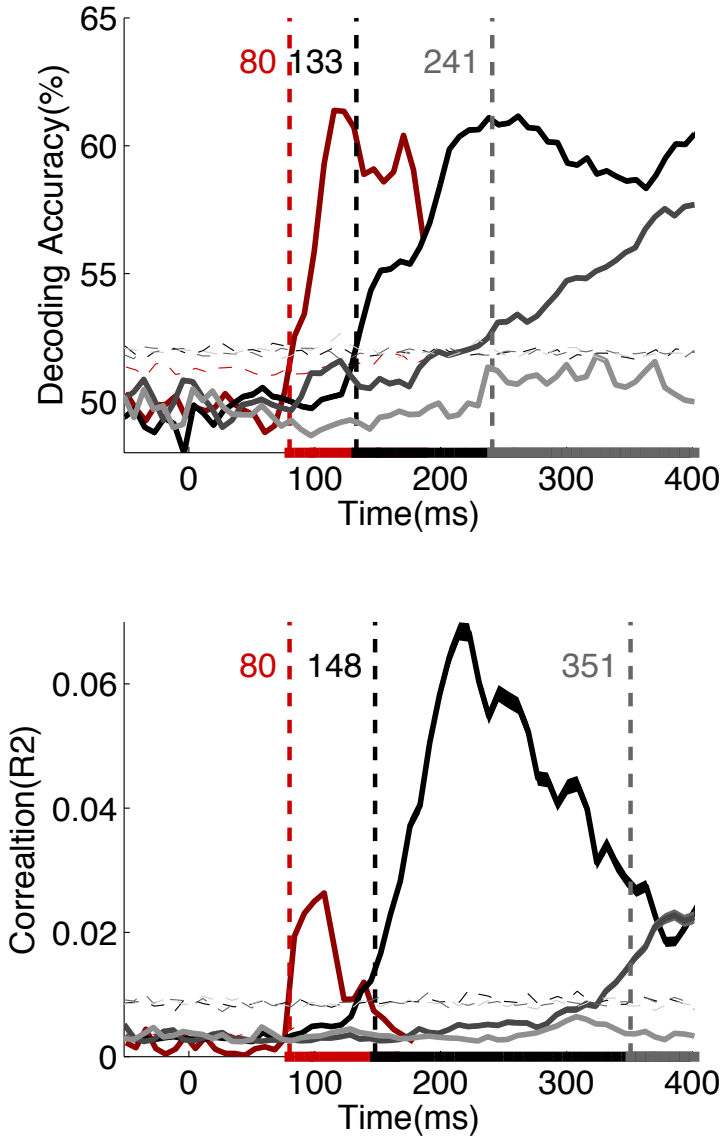

Figure 16: Décodage neuronal temporel et lien avec le comportement essai par essai. (A) Précision de décodage moyenné à travers les sujets pour les aires intermédiaires du singe (rouge) et pour les trois régions corticales d'intérêt des patients épileptiques (noir : LOAnA, gris foncé : temporale, gris clair : fronto-pariétale). Les courbes pointillées indiquent la limite supérieure de l'intervalle à 95% obtenu en randomisant les labels à travers 300 cross-validations. Les lignes verticales en pointillés indiquent la première précision de décodage significativement au dessus de la chance et suivie par 5 autres pas de temps significatifs. Les valeurs de précisons significatives ($p<0.05$) sont indiquées sur l'axe des abscisses. (B) Corrélation entre l'animalité neuronale (sortie du classifieur résultant de la procédure de décodage) et le temps de réaction à travers les essais cibles réussis. Les codes de couleurs et de significativité sont les même que dans (A).

Pour relier ces représentations neuronales au comportement essai par essai, à chaque pas de temps, on corrèle les sorties du classifieur (valeur de décision) avec les temps de réaction à travers les essais cibles corrects. La significativité de cette mesure est évaluée comme celle de la puissance de décodage, grâce à une procédure de permutation. A partir du moment où les représentations issues de la LOAnA, comme celles issues des aires intermédiaires, sont significativement décodables, elles sont pratiquement instantanément liées à la décision catégorielle du sujet (Figure 16 B). En revanche, on n'observe une corrélation entre activité temporale et comportement que 110 ms après décodage significatif.

Ensemble, ces résultats suggèrent que les aires visuelles intermédiaires, chez le singe, et les aires occipito-latérales sélectives pour les animaux, chez l'homme, jouent un rôle homologue dans la capacité de catégorisation rapide. Elles permettent un traitement précoce de la catégorie qui est directement liée aux comportements des sujets. De plus, les latences de décodage sont très proches de celles que l'on attendait, basées sur l'analyse des temps de réaction minimaux manuels et en saccade. Enfin, on retrouve, entre les deux espèces, un décalage de 50 ms entre les traitements neuronaux comme entre les temps de réaction minimaux.

Discussion

N'oublions pas l'organe qui supporte la fonction. La cognition comparée ne peut que bénéficier des avancées actuelles en neurosciences et en particulier en neuroimagerie. Montrer qu'une fonction cognitive, partagée par deux espèces, est sous-tendue par un système et des mécanismes neuronaux similaires renforce l'hypothèse d'une homologie cognitive.

Les résultats préliminaires présentés ici suggèrent l'existence de représentations neuronales communes, entre l'homme et le singe, formées dans les aires corticales visuelles intermédiaires chez le singe (V4/PIT) et dans les aires occipito-latérales sélectives aux animaux chez l'homme. Ces représentations sont corrélées avec les temps de catégorisation comportementaux des sujets, suggérant leur implication causale dans la tâche de catégorisation rapide, c'est-à-dire dans la formation de concepts visuels rudimentaires. Le décalage de 50 ms observé ici dans le traitement neuronal de la catégorie entre l'homme et le singe, fait écho à une même différence

entre le temps minimal pour saccader vers des animaux chez le singe (Girard et al. 2008) et l'homme (Kirchner et Thorpe 2005) et que l'on retrouve également dans les tâches manuelles dans notre article 1. Les latences de décodage chez les deux primates sont aussi celles prédites par les temps de réaction minimaux dans ces différentes études.

L'analyse des potentiels de champs intracrâniens semble être un formidable outil pour découvrir des homologies neurophysiologiques fonctionnelles. On enregistre l'activité électrique résultant directement du fonctionnement de populations de neurones. De plus, la qualité et la précision temporelle du signal nous permettent d'extraire un maximum d'information essai par essai et de la corréler au comportement de manière dynamique. Néanmoins, les implantations étant différentes pour chaque patient, elles permettent difficilement d'obtenir des résultats reproductibles ; l'idéal serait d'utiliser et de combiner des cohortes de plusieurs dizaines de patients.

L'IRM fonctionnelle, permettant d'imager aussi bien le cerveau humain que celui du singe, est un bon compromis pour la recherche d'homologie fonctionnelle, notamment de par l'activité « cerveau-entier » qu'elle prodigue. Une approche particulièrement prometteuse semble être la combinaison de l'IRMf chez homme et de l'IRMf plus de l'activité unitaire chez le singe. L'intérêt est d'utiliser l'IRMf pour localiser grossièrement des régions dont le fonctionnement serait similaire puis, grâce à l'enregistrement unitaire, d'analyser le véritable code neuronal. Même si cette combinaison n'a pas encore été utilisée dans une seule et même étude, à ma connaissance, plusieurs résultats récents laissent imaginer la validité et le potentiel d'une telle approche. On se contentera ici de parler des travaux relatifs au sujet qui nous intéresse dans cette thèse, c'est-à-dire la représentation des objets et catégories dans la voie visuelle ventrale. A noter cependant que les recherches sur la perception 3D ainsi que sur la manipulation d'objets et d'outils principalement par les équipes de Guy Orban et Wim Vanduffel à Leuven en Belgique et de Giacomo Rizollati à Parme en Italie, ont révélé des homologies fonctionnelles, mais également certaines différences, entre homme et macaques, principalement au sein du cortex pariétal (voir Orban 2011 pour une revue).

On connaissait l'existence d'une région localisée dans la partie ventrale du lobe temporal dédiée au traitement des visages chez l'homme (Kanwisher et al. 1997). Tsao et al. (2003) révèlent en IRMf qu'une telle région existe aussi chez le macaque rhesus. Dans une étude ultérieure, la même équipe montre que cette région, délimitée par l'IRMf chez le singe, contient effectivement une grande majorité de cellules sélectives aux visages (Tsao et al. 2006), offrant ainsi un pont entre deux champs de recherches énormes en neurosciences que sont : les corrélats neuronaux obtenus par IRMf chez l'homme et l'activité unitaire enregistré chez le singe. Un pont qui est emprunté et consolidé par une étude plus récente (Kriegeskorte et al. 2008). En utilisant cette fois un plus large panel de stimuli (visages humains, animaux, corps, objets…) Kriegeskorte et al. (2008) mettent en relation directe le traitement catégoriel des objets chez l'homme dans LOC, reflété par l'activation BOLD en IRMf, et les réponses catégorielles de populations de neurones, enregistrées en unitaire chez le macaque rhésus.

Deux autres études documentent un peu plus l'organisation similaire entre macaques et humains de ces patches sélectifs aux visages, en enregistrant dans les mêmes conditions les deux espèces (Tsao et al. 2008, Rajimehr et al. 2008). Il existerait ainsi 6 patches chez le singe, contre 4 à 5 patches chez l'homme, répartis de la même manière selon un axe occipito-temporal, avec cependant une tendance à la « ventralisation » de ces aires chez l'homme. Plus récemment Nasr et al. (2011) ont montré que cette organisation homologue pour les traitements visuels, ne se limite pas aux visages mais existe aussi pour la perception de scènes.

Une nouvelle méthode, qui vise à corréler directement le signal BOLD enregistré chez l'homme et le singe observant la même séquence vidéo, vient cependant apporter une vision légèrement différente (Mantini et al. 2012). Alors que les premières aires visuelles (V1/V2/V3) effectuent des traitements similaires, les choses semblent se passer différemment quand on monte dans la hiérarchie visuelle. Les plus fortes corrélations fonctionnelles ne sont pas forcément celles prédites par l'architecture du cerveau en suivant un modèle simple d'expansion des territoires corticaux. A travers l'évolution, il aurait pu se produire un déplacement topologique de certaines fonctions compatible avec l'hypothèse de recyclage neuronal (Anderson 2010).

Pour conclure, si une majorité d'études semble suggérer un traitement des objets et des catégories visuelles homologues dans les voies ventrales de l'homme et du singe, d'autres suggèrent certaines différences fondamentales. Le sujet est délicat et la validation de réels traitements neuronaux homologues nécessite la combinaison de nombreuses études à différents niveaux d'intégration. L'approche défendue ici, en proposant pour la première fois l'analyse conjointe de potentiels de champs intracrâniens chez l'homme et le macaque, apporte sans nul doute une petite pierre à l'édifice et permettra probablement d'initier une nouvelle approche pour la recherche d'homologies neuronales et cognitives entre ces deux primates. Le lien avec le comportement trop souvent oublié dans les études citées précédemment est aussi un point fort de l'utilisation de la tâche de catégorisation rapide qui nous permet de savoir si les corrélats neuronaux enregistrés ont vraiment une utilité pour la perception.

3.8 Discussion

Nous présentons ici les aires visuelles intermédiaires et un codage purement ascendant comme suffisant pour réaliser la tâche de catégorisation rapide. Quelle est l'influence des processus réentrants? Qu'en-est-il d'une voie visuelle rapide sous corticale ?

3.8.1 Article 5: Sur la possibilité d'un traitement purement ascendant de l'information visuelle dans la voie ventrale

Problématique

Dans l'Article 3, étant donné la rapidité avec laquelle on peut décoder l'information catégorielle et le fait que ce décodage précoce soit préservé malgré la présence d'un masque rétrograde, nous argumentons pour un codage purement ascendant de la catégorie dans la voie visuelle ventrale. Cette théorie a d'abord été proposée par Thorpe et al. (1996) basée sur leur différentielle précoce en EEG et les réponses comportementales rapides, puis reprise et théorisée, entre autres, par Vanrullen (2006) et Dicarlo (2012).

L'argumentation avancée par tous ces auteurs pour expliquer le codage purement ascendant de l'information visuelle est la suivante : étant donné la rapidité des temps de réponses comportementales en catégorisation rapide et le nombre de relais par lesquels l'information doit transiter dans la voie visuelle pour construire une représentation invariante et sélective de l'objet, le temps passé à chaque relais ne peut pas dépasser les 10 ms. En 10 ms, un neurone a uniquement le temps d'émettre 1 potentiel d'action pour transmettre l'information au relais suivant, plus haut dans la hiérarchie (Figure 17). Nos résultats confirment cette hypothèse en montrant que dans V4/PIT l'information est sélective à la catégorie dès 70 ms (c'est-à-dire aussitôt que l'information visuelle arrive dans PIT, Figure 17) et que cette latence de sélectivité influence directement les réponses comportementales des sujets.

Néanmoins, les études anatomiques, visant à retracer la connectivité entre les aires visuelles, révèlent qu'il existe un système de connexions descendantes (des aires haut-niveau vers les aires bas-niveau, Clavagnier et al. 2004) au moins aussi dense que celui des connexions ascendantes (Felleman et Van Essen 1991). Quel est le rôle de

ces connexions descendantes pour la vision et peuvent-elles intervenir dans la catégorisation rapide d'objets ? C'est la question que nous adressons dans l'Article 5 à travers le commentaire d'un article publié par Koivisto et al. (2011) dans la revue Journal of Neuroscience.

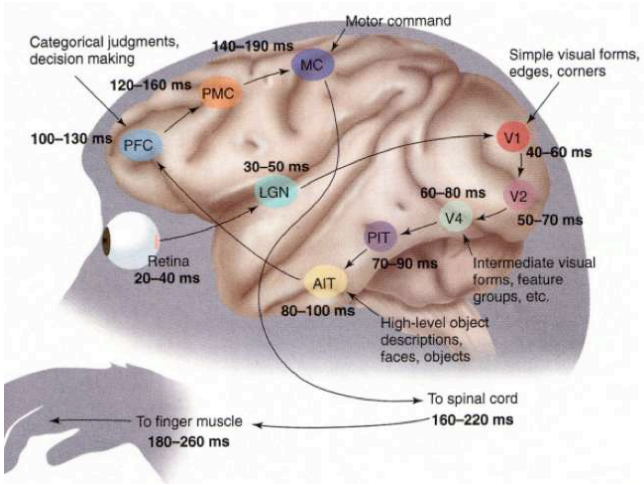

Figure 17: Latences d'activations (la plus courte et moyenne) des différentes aires impliquées dans la catégorisation d'objets (Thorpe et Fabre Thorpe 2001)

Résumé de l'article

Koivisto et al. (2011) utilisent la stimulation magnétique transcranienne (TMS) pour étudier le rôle des processus réentrants dans V1/V2 pendant la tâche de catégorisation rapide. Contrairement à ce que Thorpe et al. (1996) postulaient, ils trouvent que l'activité réentrante pourrait influencer les performances de catégorisation rapide.

Dans ce commentaire, nous essayons de réconcilier les deux approches. Notamment, nous rappelons que la notion de codage purement ascendant ne s'applique qu'aux réponses les plus rapides (voir Delorme et al. 2000). Par contre, lorsque les réponses des sujets sont déclenchées avec de plus longues latences, comme dans les cas où le stimulus demande un traitement complexe (camouflage, occlusion, etc. voir Fabre-Thorpe et al. 2001 dans le cas des cibles "difficiles"), des processus réentrants peuvent largement entrer en jeu, comme révélés par l'étude de Koivisto et al. (2011).

Enfin, nous mettons l'accent sur un résultat particulièrement intéressant de leur étude. La perturbation des processus réentrants dans V1/V2 semble avoir un effet majeur sur la capacité des sujets à juger de la qualité de leur réponse. Un résultat qui fait écho à certaines théories suggérant un rôle clé des connexions rétrogrades et des processus réentrants au niveau du cortex visuel primaire dans la conscience visuelle (Mumford 1991, Lamme et Roelfsema 2000, Bullier 2001).

ARTICLE 5: When does the visual system need to look back?

Sebastien Crouzet, Maxime Cauchoix

(2011), *Journal of Neuroscience,* 31(24), 8706-07

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

When Does the Visual System Need to Look Back?

Sébastien M. Crouzet[1] and Maxime Cauchoix[2,3]

[1]Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, Rhode Island 02912, [2]Université de Toulouse, Université Paul Sabatier, Centre de Recherche Cerveau et Cognition, 31062 Toulouse, France, and [3]Centre National de la Recherche Scientifique, Centre de Recherche Cerveau et Cognition, Faculté de Médecine de Purpan, 31062 Toulouse, France

Review of Koivisto et al.

Visual perception, and in particular object recognition, has traditionally been conceived of as a hierarchical process with a dichotomy between feedforward and feedback. Going forward in the ventral stream (V1/V2→V4→PIT→AIT), cell responses gradually become size and position tolerant, as well as selective to increasingly complex features. In this framework, the role of the early retinotopic visual cortex (i.e., V1/V2) is limited to basic computations of the visual input (e.g., edge detection) to feed higher areas for more complex processing. A single forward sweep through this hierarchy may be sufficient to perform object categorization (Serre et al., 2007). Indeed, rapid categorization experiments showed that humans and monkeys can produce extremely fast and reliable responses about the category of objects in natural scenes (Thorpe et al., 1996). It is important to keep in mind that this original hypothesis comes from a very specific observation: the first reliable behavioral responses start as early as 250 ms after image display. With roughly 10 synapses from the retina to high-level visual areas, and considering that transfer of information through each synapse takes 10–20 ms, behavioral responses occurring only 250 ms after image presentation are produced too early for recurrent processing to have a role. Nevertheless, a crucial point to note is that there is no reason to argue that later behavioral responses need to rely only on the first feedforward sweep. Late responses may take advantage of recurrent processing, for instance where the object is embedded in a cluttered background or is partially occluded, or simply when subjects are not respecting the instruction to respond as fast as possible. Indeed, a large number of feedback connections transmitting information from higher to lower levels of the hierarchy have been found anatomically in the visual system. Their functional role, so-called recurrent processing, is unaccounted for in this framework.

A recent study by Koivisto et al. (2011), published in *The Journal of Neuroscience*, set out to directly test whether recurrent processing was needed for natural scene categorization. The authors used transcranial magnetic stimulation (TMS) in combination with a rapid animal versus non-animal categorization task, a paradigm repeatedly used by Thorpe and colleagues as evidence for the role of feedforward processing for scene perception (e.g., Thorpe et al., 1996). Koivisto et al. demonstrated that V1/V2 activity continues to have an impact on the speed of categorization after the first feedforward sweep is over, which suggests that recurrent processing and/or sustained activity involving these early areas may be involved in the process of natural scene categorization. More specifically, a TMS pulse on V1/V2 with a 90 ms stimulus onset asynchrony (SOA = delay between image presentation and TMS pulse) significantly delayed the reaction time (RT), whereas a TMS pulse applied on the lateral occipital area (LO) did not affect RT for SOAs <150 ms; this is consistent with a sequential transfer of information from lower levels (V1/V2) to higher levels (LO) of the hierarchy. Somewhat less expected, pulses applied on V1/V2 at SOAs >90 ms also induced a delay in RT, as though activity in V1/V2 remained functionally important for categorization once the first round of computations was over.

Incidentally, in the Koivisto et al. study, median reaction times were far longer (between 550 and 600 ms) than those generally observed in rapid natural scene categorization tasks (between 350 and 450 ms). Thus, the responses that they recorded and analyzed were late responses, where an influence of recurrent processing might be expected, and was found (effect of a late TMS pulse on V1/V2). In any case, even if an effect was found on reaction times, the absence of any effect on accuracy suggests only a marginal role for recurrent processing in the categorization task itself.

An additional aspect of the Koivisto et al. study is that participants had to make a confidence rating for their responses (from 0 for a guess to 2 for a confident decision). These ratings can be seen as a quantitative measure of the subjective visibility. As expected, TMS made subjects less confident in their responses, possibly inducing a form of backward masking re-

Received April 14, 2011; revised May 12, 2011; accepted May 13, 2011.

Correspondence should be addressed to Sébastien M. Crouzet, Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Box 1978, Providence, RI 02912. E-mail: sebastien_crouzet@brown.edu.

DOI:10.1523/JNEUROSCI.1878-11.2011

Copyright © 2011 the authors 0270-6474/11/318706-02$15.00/0

ducing subjective visibility, but not strong enough to significantly disturb the accuracy of the response. TMS of V1/V2 affected confidence at shorter SOAs than TMS of LO, which is compatible with a disruption of the first feedforward sweep. More interestingly, V1/V2 stimulation resulted in a larger disturbance of subjective visibility than LO stimulation, and the effect of TMS on V1/V2 was still seen at SOAs longer than for LO. While the effect size is rather small (maximum drop of ~7% in the rating value at 120 ms), it is significant and there is a correlation between the level of suppression and the intensity of the pulse (illustrated in the paper at a SOA of 120 ms). All in all, this suggests that the activity in the primary visual cortex plays a functional role in subjective visibility, even after activity in LO has ceased being important. This activity could be caused by feedback from higher areas as well as by sustained activity in the early visual cortex. The sustained activity explanation might be supported by the single long effect of TMS observed in this study, in comparison to the two discrete periods reported in Camprodon et al. (2010). As a conclusion, even though this study failed to provide convincing evidence for the involvement of recurrent processing in fast categorization, it experimentally establishes a crucial link between recurrent/sustained activity in primary visual cortex and visual awareness.

The Koivisto et al. study is one of the rare experimental demonstrations in support of the theoretical works claiming an active role for V1 in visual awareness. Mumford (1991) was the first to suggest that a low-level visual area (the thalamus in his case) could be the center for visual awareness, functioning as an active blackboard for cortical reentrant activity. During the first feedforward pass, the visual signal is routed to different cortical areas for specific analysis by "experts." Then, the massive feedback projection allows these experts to "write back" the results of their analyses on the thalamic blackboard, whose role will be to bind them and generate visual awareness. The lack of direct proof of such a role for the thalamus, as well as converging evidence for a similar mechanism in primary visual cortex, has led several authors to transfer the blackboard location to V1 (Lamme and Roelfsema, 2000; Bullier, 2001). However, even if Lamme and Roelfsema (2000) have argued that feedback activity was sufficient to give rise to awareness, the mechanisms by which the reinjection from high- to low-level visual areas could give rise to awareness are unknown.

Moreover, until recently, convincing experimental evidence for a functional role of feedback to early visual areas came only from the dorsal stream, more specifically from V1/MT interactions. Using TMS, it has been shown that activity in V1 after a TMS pulse over MT was critical for conscious perception of a moving phosphene (Pascual-Leone and Walsh, 2001). Hupé et al. (1998) showed that reversible inactivation of MT in anesthetized monkeys rapidly modulates the activity of V1 cells depending on figure–ground information. The authors hypothesized that this fast feedback was made possible by the fast-conducting axons of the magnocellular pathway, which could get information to the top of the hierarchy and reinject it just in time for the arrival of the parvocellular information. Interestingly, the magnocellular path does not project only to the dorsal stream, so there is no a priori reason to think that rapid feedback as observed by Hupé et al. could not be found in the ventral stream. However, experimental evidence for the role of reentrant activity to V1 in supporting conscious object recognition has been lacking. The Koivisto et al. study, together with a paper by Camprodon et al. (2010), are the first experimental attempts to test the implication of recurrent processing in the ventral stream for conscious object recognition. Now that microstimulation in monkeys has encountered a large success to document top-down effects reminiscent of attentional modulation (Moore and Armstrong, 2003), we hope that these encouraging results using TMS will trigger more experimental investigations to elucidate the functional role of feedback.

References

Bullier J (2001) Integrated model of visual processing. Brain Res Rev 36:96–107.

Camprodon JA, Zohary E, Brodbeck V, Pascual-Leone A (2010) Two phases of V1 activity for visual recognition of natural images. J Cogn Neurosci 22:1262–1269.

Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784–787.

Koivisto M, Railo H, Revonsuo A, Vanni S, Salminen-Vaparanta N (2011) Recurrent processing in V1/V2 contributes to categorization of natural scenes. J Neurosci 31:2488–2492.

Lamme VAF, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23:571–579.

Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation. Nature 421:370–373.

Mumford D (1991) On the computational architecture of the neocortex: I. The role of thalamocortical loop. Biol Cybern 65:135–145.

Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science 292:510–512.

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A 104:6424–6429.

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522.

3.8.2 Article 6: Sur la possibilité d'une origine sous-corticale de la vision rapide

Problématique

Dans des expériences de conditionnement auditif à la peur, menées chez le rongeur, Ledoux (1996) a mis en évidence une voie sous-corticale ultrarapide permettant d'activer l'amygdale et de déclencher une réponse de peur avant même que le cortex ne soit activé. La convergence entre ces résultats chez le rongeur et la mise en évidence du rôle de l'amygdale dans le traitement des émotions, en particulier de la peur, chez l'homme et le singe a mené certains auteurs à proposer l'existence d'un circuit sous-cortical de la peur rapide, inconsciente et automatique chez le primate (Ohman et Mineka 2001), circuit qui pourrait aussi être activé par des stimulations visuelles (Vuillemier 2005). En généralisant, pourrait-on expliquer les réponses comportementales de catégorisations rapides par une telle voie sous-corticale ?

Résumé de l'article

Dans cet article, à travers une revue choisie de la littérature, nous défendons l'idée qu'il est peu probable qu'une voie sous-corticale rapide puisse être porteuse d'un signal visuel haut-niveau permettant la catégorisation rapide. La reconnaissance d'objets nécessite deux propriétés fondamentales, la sélectivité et l'invariance. Si des modèles de la voie corticale ventrale permettent d'expliquer l'émergence de ces propriétés à travers un codage hiérarchique ascendant, de tels modèles pour une voie sous-corticale sont inexistants. Ainsi la forte invariance et sélectivité de certains neurones enregistrés dans l'amygdale résultent vraisemblablement du travail de la voie corticale visuelle ventrale. De plus, alors que les expériences de catégorisation rapide témoignent d'un traitement ultrarapide de l'information, il n'existe pas de preuves comportementales convaincantes d'un traitement émotionnel réellement rapide chez le primate. Enfin, en regroupant les résultats d'études électrophysiologiques ayant rapporté des latences d'activation et de sélectivité, on se rend compte que l'activité dans la voie ventrale est compatible avec les réponses comportementales les plus rapides observées en saccade, alors que l'activité enregistrée dans l'amygdale est trop tardive. Pour conclure, nous suggérons un nouveau point de vue sur le rôle de l'amygdale dans la perception visuelle.

ARTICLE 6: How plausible is a subcortical account of rapid visual recognition?

Maxime Cauchoix, Sebastien Crouzet

En préparation

How plausible is a subcortical account of rapid visual recognition?

Maxime Cauchoix[1,2], Sébastien M. Crouzet[3,4]

[1] Centre de Recherche Cerveau et Cognition, Université Paul Sabatier, Université de Toulouse, Toulouse, France
[2] Faculté de Médecine de Purpan, CNRS, UMR 5549, Toulouse, France
[3] Cognitive, Linguistic and Psychological Science Department, Brown University, USA
[4] Institute of Medical Psychology, Charité University Medicine, Berlin, Germany

Primates are extremely fast at recognizing objects in natural visual scenes. The ventral visual cortex is usually assumed to play a major role in this ability ("high-road"). However, an alternative which is often mentioned, especially for biologically relevant or emotional stimuli, is that the visual cortex is bypassed by a rapid subcortical route to the amygdala ("low-road"). Here, we highlight the lack of evidence from psychophysics and computational models to support this low-road alternative. More, the timing of neural responses invites to seriously reconsider the role of the low-road in rapid visual categorization of natural scenes.

The speed of sight

The rapid and accurate processing of complex visual scenes is critical for the survival of primates. The high performance of the primate visual system to process complex natural scenes has been demonstrated experimentally using rapid visual serial presentations (RSVP, [1]) where human observers were able to detect a target object embedded in an image sequence even when presented at rates as fast as 10 per second. More than the high temporal resolution demonstrated using RSVP, the absolute speed of processing was demonstrated by showing that a behavioral motor response (manual or saccadic) can also be executed extremely rapidly based on visual processing. In a series of pioneering experiments, Thorpe and colleagues introduced the rapid visual categorization protocol [2-4], in which participants were able to report the presence of animals in natural scenes within only a few hundreds of milliseconds. This result set strong time constraints on the neural mechanisms underlying object categorization. But several studies went even further by showing that diagnostic category information might be readily available even earlier, considering that selective eye movement responses can be produced within only 100-120 ms of stimulus presentation [5-7]. What neural mechanisms could account for such rapid vision?

The cortical "high-road"

A widely held view about the neural mechanisms underlying object recognition is that it proceeds through hierarchically organized areas along the ventral visual stream (see [8] for a review) running from the primary visual cortex (V1) through extrastriate visual areas (V2 and V4), to the inferotemporal cortex (IT) where high-level visual representations are processed. To reconcile this view with the short behavioral latencies observed in rapid categorization tasks, several authors have suggested that a pure feedforward sweep of activity through the ventral stream might be sufficient to perform core object recognition [8-10].

The subcortical "low-road"

On the other hand, a subcortical shortcut, the so-called "low-road", might also be a plausible alternative. This hypothesis finds its origin in the rapid amygdala responses reported by Ledoux during auditory fear conditioning [11]. In a series of experiments in rodents, he delineated a quick route which bypasses the cortex by directly reaching the amygdala via the thalamus. Such a subcortical shortcut would allow, in specific cases such as threat, to quickly initiate appropriate defense responses even before the sensory cortices become involved. Further, once amygdala was also reported to be a hub for emotion (particularly fear) recognition in humans [12], this alternative pathway was proposed to account for rapid, automatic and unconscious reactions among humans and monkeys to biologically relevant visual stimuli [13-19].

Here we argue that the role of this low-road, when extended to rapid visual object processing in primate, does not seem to be supported by convincing computational and behavioral evidence (1. and 2.). To preface our arguments, first, real-world object categorization requires computational properties that have not yet been found in a subcortical pathway. Second, among the characteristics attributed to this low-road, we argue that genuine rapidity has not yet been demonstrated appropriately. Finally, the low-road hypothesis is at odds with neural latencies reported in the amygdala and the visual cortex (3.). This clearly point to an earlier role for cortical areas and suggests a serious reconsideration of the role of the "low-road" in rapid vision.

1. Real-world recognition requires selectivity and invariance

Real-world recognition of objects (even a fearful face or a snake) is a hard problem. It requires both invariance and selectivity. Each object can be encountered at any location on the retina (position variability), at a range of distances (scale variability), at many angles relative to the observer (pose variability), at a range of lighting conditions (illumination variability) and in different visual contexts (clutter variability). To support recognition a neural system needs to deal with the inherent variability of sensory input and solve the so-called "invariance problem" [8]. An opposite demand for a recognition system is the need to reach a high level of selectivity [20]. For example in the facial emotion case, discriminating between fear and surprise, or disgust and anger certainly requires sensitivity to very modest variations. The balance between invariance and selectivity is a hallmark feature of visual recognition in primates (and still a formidable challenge for computer vision).

In macaque monkeys neural responses selective and invariant to complex objects are typically found in the inferotemporal cortex (IT; see [8] for a review). Neural responses in this area are tolerant to retinal position, scale or pose of the object [21]. Studies using intracranial recordings in human epileptic patients have also shown that neural responses from the visual cortex provide a categorical signal tolerant to changes in scale and position [22].

Driven by results from electrophysiology, a plausible model of how selectivity and invariance could be built through the ventral stream has emerged. It is based on two successive simple operations (template-matching and nonlinear pooling) repeated at each stage of the ventral hierarchy [8,10,20,23]. Such hierarchical models have been shown to accurately mimic primate rapid categorization performance [10,24] and neural responses of the visual ventral stream [25,26].

Among subcortical structures, human single-unit studies showed that the amygdala contains neurons selective to categories or

objects such as animals, famous faces or places [27-29]. Interestingly, these neurons are highly invariant since they respond to various pictures of their preferential objects, but also to their written or spoken names. However, there is currently no model of how this high level of both selectivity and invariance could be built from a direct subcortical route. A more reasonable assumption would thus be that it gets its input from the ventral stream (IT / LOC), rather than from the thalamus (shortcut "low-road" route).

2. What is rapid visual processing?

Numerous studies investigating affective stimulus processing have used short image presentation and/or masking protocols to show that emotions such as fear can be processed unconsciously and "rapidly" [30-32]. There is no doubt that masking is a powerful tool for revealing unconscious sensory processing. However, this experimental tool cannot inform about the genuine rapidity of visual processing. In backward masking protocols, the stimulus onset asynchrony (SOA, time interval between target and mask onset) is a measure of the visual uptake time (or temporal resolution), not of the time required for complete visual processing. In other words, even in a perfect pipeline model of the visual system, the mask interference would only give information about the time spent at each stage, and not about the cumulative time for all stages [33]. For example, the fact that fear information can be extracted from faces masked after an SOA of 39 ms [30] informs us about the minimal visual uptake time necessary for fear processing but does not say anything about the time at which fear information is available to trigger behavioral responses.

The speed of processing for object or scene

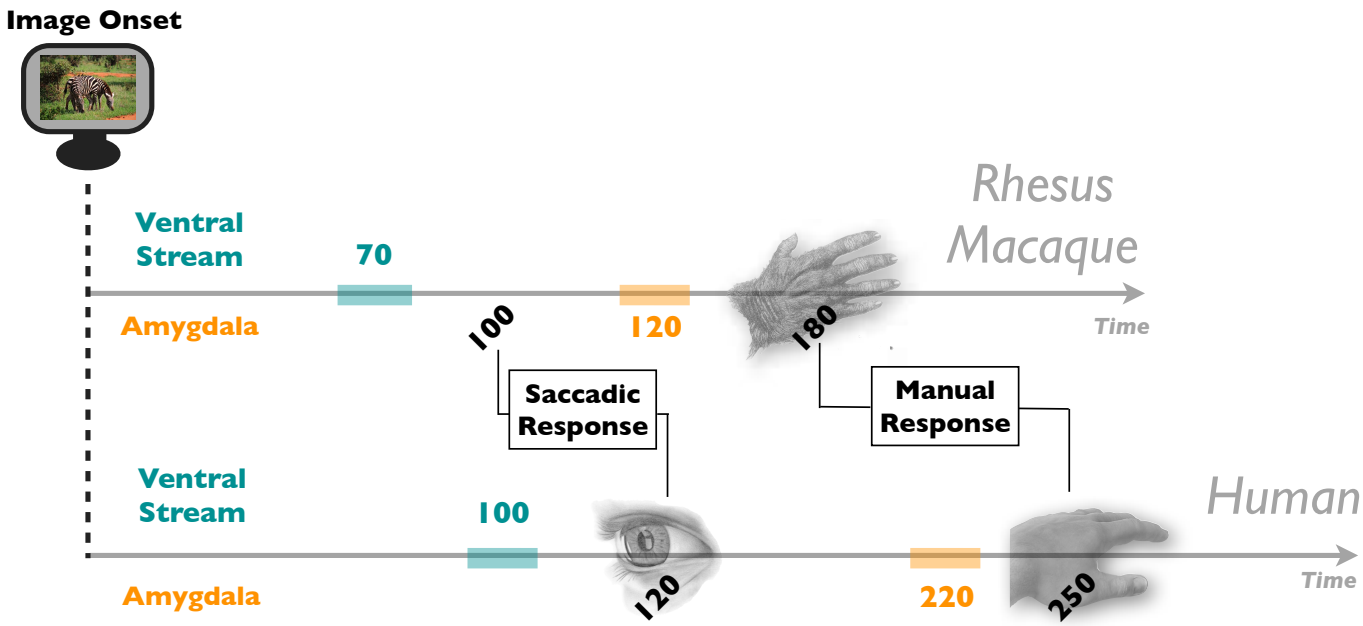

Figure 1: Fastest neural and behavioral visual responses to the animal category

Based on a review of the existing literature, we report here the fastest behavioural responses among monkeys (top) and humans (bottom) in a rapid animal/non animal categorization task. The fastest reaction times reported in studies using a saccadic forced-choice [5,6] are represented with an eye picture, those using a manual go/no-go task [2,3] with a hand picture. Colored areas correspond to the earliest known latency of neurons selective to the animal category reported in the ventral stream (turquoise; Cauchoix et al., submitted for monkeys; [22] for humans) or in the amygdala (orange; [29,40] for humans; [38] for monkeys). The neural latencies reported here were either the earliest selective neural activity [22,26,38] or the average latency of selective neurons [29,40]. In any case, when taken together, they strikingly suggest that the ventral stream is much more likely to contribute to rapid vision than the amygdala.

categorization has been extensively studied using rapid categorization protocols. Using minimal reaction time measurements (time after stimulus onset at which responses start to be reliable: when correct responses start to significantly outnumber incorrect ones) it has been shown that humans can categorize images as containing an animal in only 250 ms [2], while monkeys can perform the same task by 180 ms [3]. Even faster, reliable saccades towards faces and animals can be triggered as soon as 100-120 ms after image onset [5-7]. As far as we know, there is no evidence for faster processing of emotional stimuli, as would be predicted by the low road hypothesis.

3. Ventral stream visual cortex is activated before the amygdala

Most of the studies on humans investigating the role of the amygdala in visual processing have used fMRI and PET scans ([34,35], see [15,36] for reviews). These two techniques, because of their poor temporal resolution, do not allow conclusions about the temporal dynamics of stimulus processing. Despite this limitation, it was assumed that amygdala responses to emotional stimuli, notably to fearful stimuli, were based on a rapid low-road activation [13,15].

A review of electrophysiological studies reporting neural latencies suggests a clearly different picture (Fig. 1). Recent investigations in both multi-unit [21] and intracranial field potential (IFP, Cauchoix et al., submitted) showed that object category information can be read out from the monkey visual ventral stream as early as 70-80 ms after stimulus onset. Similarly, in human epileptic patients, IFP recorded from the occipito-temporal cortex were selective to object category as early as 100 ms after stimulus onset [22]. These category selective latencies are compatible with the rapid behavioral responses observed in natural scene categorization tasks [2,3,5,6]. Furthermore, the early ventral stream activity seems to be directly linked with rapid behavioral responses (Box 1). On the contrary, selective responses to visual features or objects in the amygdala tend to occur later for both monkeys [37,38] and humans [39,40]. In addition, one single-unit study [41] performed a direct comparison (on the same monkeys) between the two pathways and showed that neurons in the STS (top of the ventral stream) had latencies (90-140 ms) that systematically preceded those observed in the amygdala (110-200 ms). The pattern observed thus clearly vouches for a cortical "high-road" precedence.

BOX 1: Ventral stream activity and object recognition

Impairment of visual identification and discrimination after lesion of the ventral stream was the first direct evidence for a role of the occipitotemporal network in object recognition [42]. Recently, the dynamics of the ventral stream contributions to visual perception have been more precisely documented. The stimulation of IT face selective patches 50 to 100 ms after image presentation induced a strong bias in the monkeys to categorize an image as containing a face [43]. Similar results have been obtained in humans using early (60-100 ms after stimulus onset) transcranial magnetic stimulation on LOC. Such stimulations impaired behavioural responses in visual discrimination tasks [44] and modulated EEG signals of early selective activity [45]. Finally, in monkeys, early selectivity (<100ms) in the ventral stream seems to directly drive behavioral responses in the rapid categorization task [26]. Altogether, these studies support a very early link (<100ms) between ventral stream activity and object recognition.

Conclusion

In this paper we questioned the hypothesis that a subcortical low-road could account for the speed of sight. Several observations from psychophysics, computational modeling and electrophysiology strongly suggest that the low-road account is mostly incompatible with the characteristics of rapid visual categorization. On the contrary, a large collection of evidence confirms that the cortical high-road, through the visual ventral stream, can accomplish a rapid, selective and invariant analysis of the scene. The latency of neural visual activation and response characteristics in the amygdala clearly suggest that its involvement in visual processing is downstream of the ventral visual cortex, after core object recognition has been performed. The exact role of amygdala's late contribution to visual perception remains an open question (see [36] and Box 2 for specific speculations).

BOX 2: What is the role of amygdala in visual perception?

Numerous studies and observations suggest a role for the amygdala in a reentrant processing framework of visual perception [46-49]. First, a rapid feedforward sweep through the ventral stream would produce a coarse/basic visual representation. This would be sufficient to enable rapid behavioral responses, such as those observed during rapid categorization [8], but also rapid and unconscious emotion recognition even when the amygdala is lesioned [50].

The visual information would then be further processed by "expert" areas to enrich this basic representation through reentrant processing. In this framework, the amygdala would play the role of an emotional expert, coloring visual representations with their emotional valences. Indeed, anatomical studies revealed numerous connections from amygdala to visual cortical areas [51]. A functional aspect of this communication between the amygdala and the visual cortex has been illustrated by recording fMRI responses to emotional stimuli in both patients with amygdala lesion and control subjects [52]. The presentation of fearful faces increased activation of fusiform and occipital cortex in controls, but not in patients with damage to the amygdala. Electrophysiological recordings demonstrated a modulation of responses by the emotional content of stimuli. This modulation appeared earlier in the amygdala than in visual cortical areas [39]. A complementary observation is that the initial response of IT neurons was shown to carry information about the global category of objects (such as human or monkey face), while the modulation by the emotional content of the stimuli (such as facial expression) appeared later (by about 50 ms, [53]). This interval of time fits perfectly with a quick round-trip between IT and the amygdala. At the phenomenological level, [54] reported the case of a patient with a lesion of the connection between the amygdala and the visual cortex (Capgras syndrome). This patient was able to recognize the face of his mother, but could not link it to an appropriate affective judgement, and hence considered his mother to be an imposter. This could be a manifestation of a lack of recurrent processing between the amygdala and the visual cortex.

The conscious visual perception with all its richness would finally emerge from the interplay between the "expert" and the retinotopic "blackboard" [48], explaining why the amygdala seems necessary for the conscious experience of fear [50,55].

Outstanding questions

1 - Latencies reported in the literature usually correspond to the start of a neural response following visual presentation, or the average latency of neurons selective for a category of interest. Unfortunately, few studies have looked for the selectivity latency, i.e. the time at which neurons or LFP/IFP become selective to visual features. Single-trial multivariate pattern analysis, which can be applied both to single-unit [21] and IFP/LFP [22,26], as well as information theory methods [53], seems to be a promising tool to determine selectivity latency. An effort should be made to use standard methods and statistical tools and systematically report the selectivity latencies.

2 - Few studies have used intracranial field potential in human epileptic patients to assess amygdala activity. However, because of its particular organization where neurons are multipolar with dendrites passing in all directions [56], local synaptic currents are likely to cancel each other rather than spatially summate. Thus, large potentials are unlikely to be generated in the amygdala [57] but rather originate from the entorhinal and perirhinal cortices which lie immediately inferior and medial to the human amygdala. Further experiments, using more accurate recordings, will be needed to study the functional role of the amygdala in visual processing.

3 - How rapidly can visual emotional valence be processed? Can it be faster than object recognition? More behavioral investigation using rapid categorization protocols are needed.

4 - If visual responses in primate amygdala are slow, other subcortical structures show neurons with very short visual latencies (pulvinar, superior colliculus). Do these fast subcortical structures show selective and invariant responses to complex visual objects/categories?

5 - Where does the visual information found in the amygdala come from? Neural latencies suggest the ventral stream, but more evidence is needed, for instance using microstimulation.

REFERENCES

1 Potter, M. C. and Faulconer, B. A. (1975) Time to understand pictures and words. Nature. 253, 437–438

2 Thorpe, S. et al. (1996) Speed of processing in the human visual system. Nature. 381, 520–522

3 Fabre-Thorpe, M. et al. (1998) Rapid categorization of natural images by rhesus monkeys. Neuroreport. 9, 303–308

4 Fize, D. et al. (2011) Humans and monkeys share visual representations. Proc. Natl. Acad. Sci. U.S.A. 108, 7635–7640

5 Kirchner, H. and Thorpe, S. J. (2006) Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Research. 46, 1762–1776
6 Girard, P. et al. (2008) Ultra-rapid categorisation in non-human primates. Anim Cogn. 11, 485–493
7 Crouzet, S. M. et al. (2010) Fast saccades toward faces: face detection in just 100 ms. Journal of Vision. 10, 16.1–17
8 Dicarlo, J. J. et al. (2012) How Does the Brain Solve Visual Object Recognition? Neuron. 73, 415–434
9 Vanrullen, R. and Thorpe, S. J. (2001) The time course of visual processing: from early perception to decision-making. J Cogn Neurosci. 13, 454–461
10 Serre, T. et al. (2007) A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U.S.A. 104, 6424–6429
11 LeDoux, J. (1996) The EMOTIONAL BRAIN: The Mysterious Underpinnings of Emotional Life, (1st edn) Simon & Schuster.
12 Adolphs, R. et al. (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature. 372, 669–672
13 Öhman, A. and Mineka, S. (2001) Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychol Rev. 108, 483–522
14 Öhman, A. (2005) The role of the amygdala in human fear: automatic detection of threat. Psychoneuroendocrinology. 30, 953–958
15 Vuilleumier, P. (2005) How brains beware: neural mechanisms of emotional attention. Trends in Cognitive Sciences. 9, 585–594
16 de Gelder, B. et al. (2011) Emotion in the brain: of low roads, high roads and roads less travelled. Nat Rev Neurosci. 12, 425–author reply 425
17 Tamietto, M. and de Gelder, B. (2010) Neural bases of the non-conscious perception of emotional signals. Nat Rev Neurosci. 11, 697–709
18 Tamietto, M. et al. (2009) Unseen facial and bodily expressions trigger fast emotional reactions. Proc. Natl. Acad. Sci. U.S.A. 106, 17661–17666
19 Johnson, M. H. (2005) Subcortical face processing. Nat Rev Neurosci. 6, 766–774
20 Riesenhuber, M. and Poggio, T. A. (1999) Hierarchical models of object recognition in cortex. Nat Neurosci. 2, 1019–1025
21 Hung, C. et al. (2005) Fast readout of object identity from macaque inferior temporal cortex. Science. 310, 863–866
22 Liu, H. et al. (2009) Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron. 62, 281–290
23 Fukushima, K. (1980) Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 36, 193–202
24 Crouzet, S. M. and Serre, T. (2011) What are the visual features underlying rapid object recognition? Front Psychol. 2,
25 Serre, T. et al. (2007) A quantitative theory of immediate visual recognition. Prog Brain Res. 165, 33
26 Cauchoix, M. et al. Feedforward ventral stream neural activity enables rapid visual categorization.
27 Kreiman, G. et al. (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci. 3, 946–953

28 Quiroga, R. Q. et al. (2005) Invariant visual representation by single neurons in the human brain. Nature. 435, 1102–1107
29 Mormann, F. et al. (2011) A category-specific response to animals in the right human amygdala. Nat Neurosci. 14, 1247–1249
30 Bar, M. et al. (2006) Very First Impressions. Emotion. 6, 269–278
31 Öhman, A. et al. (2007) On the unconscious subcortical origin of human fear. Physiology & behavior. 92, 180–185
32 Adolphs, R. (2008) Fear, faces, and the human amygdala. Current Opinion in Neurobiology. 18, 166–172
33 Vanrullen, R. (2011) Four common conceptual fallacies in mapping the time course of recognition. Front Psychol. 2, 365
34 Whalen, P. et al. (2004) Human Amygdala Responsivity to masked fearful eye whites. Science. 306,
35 Morris, J. S. et al. (1999) A subcortical pathway to the right amygdala mediating “unseen” fear. Proc. Natl. Acad. Sci. U.S.A. 96, 1680–1685
36 Pessoa, L. and Adolphs, R. (2010) Emotion processing and the amygdala: from a 'low road' to “many roads” of evaluating biological significance. Nat Rev Neurosci. 11, 773–783
37 Sanghera, M. K. et al. (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. Experimental neurology. 63, 610–626
38 Gothard, K. M. et al. (2007) Neural responses to facial expression and face identity in the monkey amygdala. J. Neurophysiol. 97, 1671–1683
39 Krolak-Salmon, P. et al. (2004) Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human. Neuron. 42, 665–676
40 Mormann, F. et al. (2008) Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. J Neurosci. 28, 8865–8872
41 Leonard, C. M. et al. (1985) Neurons in the amygdala of the monkey with responses selective for faces. Behav. Brain Res. 15, 159–176
42 Schiller, P. H. (1995) Effect of lesions in visual cortical area V4 on the recognition of transformed objects. Nature. 376, 342–344
43 Afraz, S.-R. S. et al. (2006) Microstimulation of inferotemporal cortex influences face categorization. Nature. 442, 692–695
44 Pitcher, D. et al. (2007) TMS evidence for the involvement of the right occipital face area in early face processing. Curr. Biol.
45 Sadeh, B. et al. (2011) Stimulation of category-selective brain areas modulates ERP to their preferred categories. Curr Biol. 21, 1894–1899
46 Mumford, D. (1991) On the computational architecture of the neocortex: I. The role of thalamo-cortical loop. Biol Cybern. 65, 135–145
47 Mumford, D. (1992) On the computational architecture of the neocortex: II The role of cortico-cortical loops. Biol Cybern. 66, 241–251
48 Bullier, J. (2001) Integrated model of visual processing. Brain Res. Brain Res. Rev. 36, 96–107
49 Crouzet, S. M. and Cauchoix, M. (2011) When does the visual system need to look back? J Neurosci. 31, 8706

50 Tsuchiya, N. et al. (2009) Intact rapid detection of fearful faces in the absence of the amygdala. Nat Neurosci. 12, 1224–1225
51 Amaral, D. G. et al. (2003) Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. Annu Rev Neurosci. 118, 1099–1120
52 Vuilleumier, P. et al. (2004) Distant influences of amygdala lesion on visual cortical activation during emotional face processing. Nat Neurosci. 7, 1271–1278
53 Sugase, Y. et al. (1999) Global and fine information coded by single neurons in the temporal visual cortex. Nature. 400, 869–873
54 Hirstein, W. and Ramachandran, V. S. (1997) Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. Proceedings. Biological sciences / The Royal Society. 264, 437–444
55 Feinstein, J. S. et al. (2011) The human amygdala and the induction and experience of fear. Curr Biol. 21, 34–38
56 McDonald, A. J. (1992) Cell types and intrinsic connections of the amygdala. In The Amygdala. Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. Aggleton, J. P., ed), pp. 67–93, Wiley-Liss
57 Halgren, E. et al. (1995) Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. Electroencephalography and clinical neurophysiology. 94, 229–250

Discussion

Dans cet article d'opinion, nous posons un regard critique sur l'existence d'une voie sous-corticale permettant une reconnaissance rapide de stimuli visuels, même émotionnels. Nos arguments fonctionnels viennent s'ajouter à ceux défendus par Pessoa et Adolphs (2010). Une de leurs critiques basée de sur la connectivité est particulièrement forte. La voie sous corticale, comme proposée par ses partisans, passe par deux relais entre la rétine et l'amygdale : les colliculus supérieurs puis le pulvinar dans le thalamus. Or, s'il existe bien une connexion entre les colliculus supérieurs et les noyaux inférieurs du pulvinar, ainsi qu'entre les noyaux médians du pulvinar et l'amygdale, il n'y en a aucune actuellement connue entre les noyaux inférieurs et médians du pulvinar. Les entrées visuelles du pulvinar, provenant des colliculus, sont donc transférées vers le cortex visuel, avec lequel les noyaux inférieurs du pulvinar échangent de nombreuses connections, et non vers l'amygdale. Ainsi, la voie sous-corticale colliculus-pulvinar-amygdale est anatomiquement inexistante chez le primate.

Dans notre article, nous mettons aussi en doute la réelle rapidité du traitement de stimuli émotionnels visuels chez le primate (notamment des expressions faciales). Cependant, on ne peut nier l'existence d'un traitement privilégié de certains types de stimuli visuels. Nous avons déjà parlé de l'expérience de New et al. (2007) démontrant que l'on perçoit plus rapidement un changement affectant un animal comparé à un objet non animé. Nous avons aussi évoqué les saccades ultrarapides et automatiques vers les visages humains rapportées par Crouzet et al (2010). A la lueur de ces résultats, je m'accorde avec les tenants de la voie sous-corticale (en particulier Ohman et Mineka 2001) sur le fait que l'évolution ait pu sélectionner des mécanismes nerveux permettant de traiter rapidement et efficacement les stimuli les plus pertinents pour la survie de notre espèce. Mais si un tel avantage ne peut pas être expliqué par un traitement sous-cortical, comme nous le défendons ici, quel pourrait en être le substrat neuronal ?

En discussion du chapitre 2, nous avons évoqué le statut particulier des serpents pour les primates. Notamment, nous avons listé certaines expériences comportementales mettant en exergue un traitement privilégié des serpents dans des tâches de recherche visuelle (Öhman et al. 2001, LoBue et DeLoache 2008, Shibasaki et Kawai 2009). Ce

statut particulier semble en partie inné puisque les singes, comme les enfants, montrent une prédisposition à acquérir la peur des serpents (Cook and Mineka 1981, DeLoabche and LoBue 2009). Les serpents, prédateurs ancestraux des primates (Isbell 2005), sont l'archétype de stimuli, pour lesquelles certains auteurs (Ohman et Mineka 2001), dans une logique évolutionniste, ont postulé l'existence d'un circuit nerveux spécifique, qui selon eux était la voie sous-corticale.

Nous avons déjà décrit les résultats préliminaires comportementaux d'une étude où des visages, des serpents et des objets neutres étaient présentés à nos deux singes. Nous avons montré que, bien qu'ils n'aient jamais eu directement à faire à de vrais serpents, les deux animaux présentaient une inhibition spontanée pour répondre en touchant les photographies de serpents. Un comportement qui est maintenu pendant les 13 jours d'expériences, chez un des deux singes, alors qu'il va à l'encontre du protocole de renforcement.

Pendant les 5 derniers jours de cette expérience, nous avons enregistrés, chez les deux singes, des potentiels de champs intracrâniens provenant de la voie visuelle ventrale. L'analyse des potentiels évoqués classiques montre clairement une spécificité du traitement des serpents par rapport aux autres catégories (
Figure 18 A). Cette différence apparaît très tôt ; environ 60 ms après la présentation du stimulus. Pour quantifier le codage de chaque catégorie, on utilise une procédure de décodage (identique à celle menée dans les autres chapitres), dans laquelle on évalue la capacité d'un classifieur à dissocier chaque catégorie d'un même jeu de distracteurs familiers. On peut décoder significativement les serpents plus tôt (65-70 ms après présentions du stimulus) que les autres catégories (80-120 ms,
Figure 18 B). La précision de décodage dans l'intervalle 50-100 ms est aussi supérieure pour les serpents (75% contre <65% pour les autres, Figure 18 C).

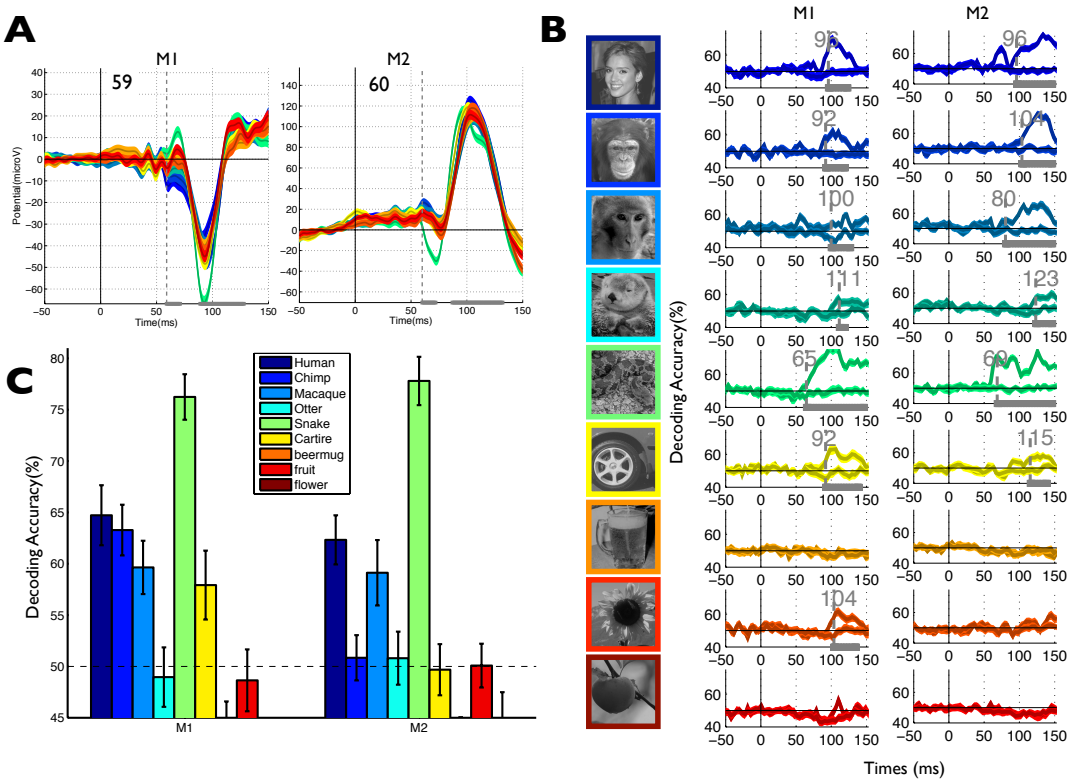

Figure 18 : Traitement neuronal spécifique des serpents dans la voie visuelle ventrale. Deux exemples d'ERPs (un pour chaque singe M1 et M2) illustrant l'effet catégoriel précoce (anova p<0.01), principalement dû aux serpents (vert) (A). Pour évaluer la quantité d'information disponible dans l'ensemble du signal, on utilise une procédure de décodage (similaire à celle reportée dans les chapitres précédents) dans laquelle on évalue la capacité d'un classifieur à dissocier chaque catégorie (couleur) des mêmes stimuli distracteurs familiers. Le décodage temporel (+/- intervalle de confiance à 95 %) nous indique quand (marque grise sur l'abscisse, début : pointillé verticaux) la précision de décodage est significativement différente d'un décodage avec labels permutés (B) Le décodage global (+/- intervalle de confiance à 95 %) en utilisant l'ensemble du signal entre 50 et 100 ms quantifie précisément la capacité du classifieur à lire chaque catégorie dans cette période (C).

Cet avantage temporel et quantitatif pour le codage des serpents s'explique-t-il par des statistiques bas-niveau des images ? On sait par exemple que le contraste global de l'image affecte les potentiels évoqués (Macé et al. 2005). Pour éviter ce type de contamination bas niveau, toutes les images (présentées en noir et blanc) ont été égalisées en luminance (moyenne des valeurs de gris à travers les pixels) et en contraste RMS (déviation standard de la distribution des niveaux de gris des pixels). Dans cette étude, on utilise aussi un nombre limité d'images par catégorie (10 exemplaires), la variabilité entre les stimuli au sein d'une catégorie est par conséquent relativement faible. Il est possible que les images de serpents soient plus similaires les unes par rapport aux autres, permettant ainsi au classifieur de généraliser plus facilement, et donc d'obtenir de meilleures performances de décodage neural. L'image moyenne est un bon moyen pour se rendre compte de la variabilité intra-

classe (Thierry et al. 2007). Alors que dans les images moyennes des visages on distingue clairement un visage moyen, l'image moyenne est beaucoup moins claire pour les serpents (Figure 19 A). Si nos résultats étaient liés à la variabilité intra-classe, on s'attendrait donc à mieux décoder les visages et non les serpents.

Enfin, pour estimer quelle aire visuelle pourrait être à l'origine de ce traitement privilégié des serpents, on utilise un modèle hiérarchique ascendant de la voie visuelle ventrale biologiquement inspiré (Serre et al. 2007). Ce modèle permet d'extraire des caractéristiques de l'image comme le ferait des neurones situés dans les différentes aires de la voie ventrale (V1, V4, IT). Pour modéliser la rétine, on utilise simplement les niveaux de gris de l'image, préalablement sous-dimensionnée en 32 * 32 pixels.

Après avoir extrait ces différentes caractéristiques des images, on utilise une procédure de classification identique à celle utilisée pour le signal neuronal, dans laquelle on demande à un classifieur de dissocier chaque catégorie d'un même jeu de distracteurs familiers. On observe, en accord avec ce qui apparaît avec les images moyennes (Figure 19 A), qu'un modèle de la rétine montre ses meilleures performances sur les visages de chimpanzés (Figure 19 B). Le traitement privilégié des serpents, lui, semble apparaître dès V1 puis disparaît dans les aires plus haut niveau (V4 et IT). C'est donc un modèle des cellules complexes de V1 qui capture le mieux les différences observées dans le décodage du signal neural (Figure 18 C). Comme notre classifieur, un neurone, qui écoute l'activité d'une population de neurones complexes de V1, pourra prédire la présence d'un serpent avec une meilleure fiabilité qu'il ne le ferait pour d'autres stimuli biologiquement pertinents comme les visages. Ces résultats sont en accord avec les latences de décodage très courtes (65-70 ms) des serpents, qui correspondent aux latences de V1 les plus tardives (Bullier et al 2001). L'avantage du codage des serpents dès V1, est que ces neurones, en ayant de petits champs récepteurs, fournissent une information précisément localisée ; une caractéristique de première importance pour la détection d'un prédateur mortel.

La simple comparaison avec un modèle n'est pas suffisante pour assurer que notre traitement privilégié des serpents provient strictement de V1. En effet, les latences de

décodage sont aussi compatibles avec les latences les plus précoces de V4 (Thorpe et Fabre-Thorpe 2001). De plus, tout comme les cellules mimant V1 du modèle de Serre et al. (2007), certains neurones de V4 répondent à une combinaison de fréquences spatiales et d'orientations (David et al. 2006). Elles peuvent aussi être sélectives à des ronds concentriques (Kobatake & Tanaka 1994), rappelant la forme d'un serpent enroulé. Et, plus généralement, elles sont sensibles à la courbure des stimuli visuels (Gallant et al. 1993, Pasupathy et Connor 1999, 2001).

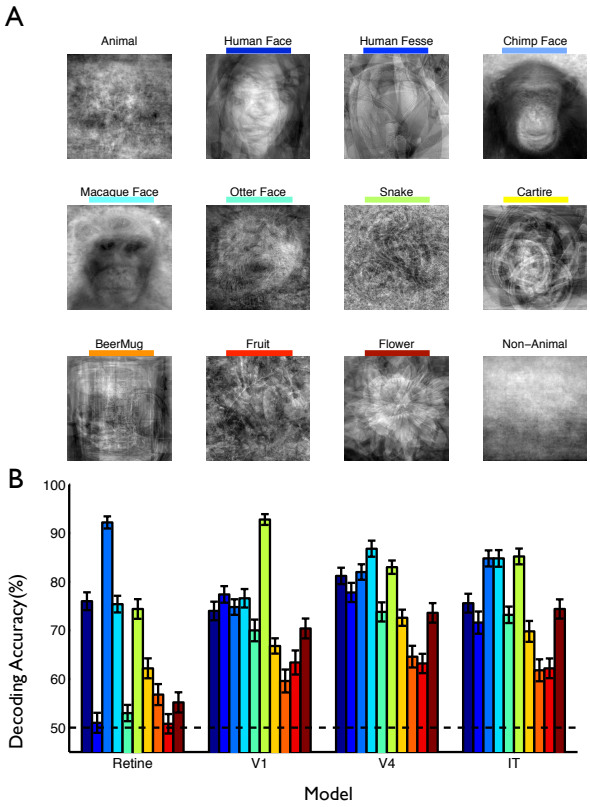

Figure 19: Statistiques d'images. Les images moyennes (moyenne des pixels entre les 10 exemplaires) fournissent une idée de la variabilité intra-classe (A). Dans une procédure similaire à celle utilisée pour le signal neuronal, on classifie les images de chaque catégorie en utilisant des caractéristiques de l'image qui correspondent à celles qu'utiliseraient différentes aires de la voie visuelle (Serre et al. 2007). La précision de décodage (+/- intervalle de confiance à 95%) est rapportée pour chaque catégorie (légende des couleurs en A) et pour chaque modèle. Le niveau de chance est indiqué par la ligne en pointillée (B).

Ensemble, ces résultats préliminaires tendent à supporter l'hypothèse selon laquelle un traitement cortical précoce dans la voie visuelle ventrale permettrait au cerveau de décider en moins d'une centaine de millisecondes si un serpent est présent dans notre champ visuel. Cette fonction est sans nulle doute un avantage majeur qui a pu être sélectionné au cours de l'évolution. Néanmoins, la réelle rapidité du traitement visuel des serpents reste encore à démontrer au niveau comportemental. La tâche de saccade

rapide développée par Kirchner et Thorpe (2005) apparaît comme le protocole idéal pour cette démonstration. Alors que le traitement des visages a été montré comme automatique et ultra-rapide dans cette tâche (Crouzet et al. 2010), il serait particulièrement intéressant d'évaluer si la perception des serpents conserve l'avantage temporel prédit par nos résultats neuronaux, par rapport à la perception des visages.

Ces données préliminaires et les idées défendues dans notre article d'opinion convergent pour montrer l'inutilité d'invoquer une voie sous-corticale pour expliquer les phénomènes de reconnaissance ou de catégorisation visuelle rapide. Même dans la mise en place d'un mécanisme visuel aussi primitif, et potentiellement pré-câblé, qu'est la détection de prédateurs ancestraux, le néocortex semble jouer un rôle prépondérant.

4 Discussion générale

4.1 Résumé des résultats

L'image de l'animal dans la société a bien changé au cours du siècle dernier. Avec l'avènement de l'éthologie et de la psychologie comparée, on a vu, petit à petit, tomber nombre de préjugés et d'idées reçues sur les animaux, au profit d'une approche scientifique et objective du comportement animal. On s'est alors rendu compte qu'il existait non pas une, mais des intelligences animales, dont celle de l'homme fait partie. Néanmoins, alors que les exemples d'études éthologiques montrant l'existence de capacités très « humaines » chez l'animal telles que, l'abstraction, l'utilisation d'outils, la planification, la métacognition ou encore l'empathie, foisonnent, peu de travaux permettent de mettre en évidence l'existence d'une origine commune de ces capacités entre l'homme et l'animal et donc de supporter l'hypothèse de continuité pour la cognition, postulée par Darwin.

Dans cette thèse nous émettons l'hypothèse selon laquelle la capacité d'abstraction, c'est à dire la capacité de former des concepts mentaux de catégories d'objets ou d'événements existant dans le monde extérieur, jouerait un rôle clé dans l'émergence de la pensée et serait partagée par de nombreuses espèces. En particulier, chez les primates, animaux dirigés par la vision, nous argumentons ici pour la continuité des mécanismes autorisant la formation de concepts visuels.

Afin d'étudier ces briques cognitives, nous avons eu recours à la tâche de catégorisation rapide de scènes naturelles. Une tâche permettant de circonscrire les mécanismes cognitifs et neuronaux suffisant pour l'élaboration de concepts visuels complexes.

Dans un premier temps, en confirmant les résultats de nos prédécesseurs, et en levant l'ambiguïté sur la possibilité d'utilisation d'indices visuels de bas-niveau par les singes, nous avons montré à quel point l'homme et le macaque semblent utiliser des représentations et des stratégies communes pour résoudre cette tâche. En engageant humains et singes dans une situation expérimentale rigoureusement identique, et en contraignant l'homme à répondre rapidement, nous fournissons un premier argument

comportemental pour l'existence d'un processus commun d'abstraction visuelle entre ces deux espèces.

Mais les avancées récentes des neurosciences nous permettent d'aller plus loin que ces simples comparaisons comportementales. En développant des méthodes d'enregistrement et d'analyses novatrices, on est capable de suivre essai par essai, l'émergence de représentations visuelles chez le singe. Plus précisément, nous démontrons comment la voie visuelle ventrale du singe permet de former très rapidement des concepts visuels remplissant toutes les caractéristiques suffisantes pour sous-tendre la capacité de catégorisation naturelle rapide de ces animaux. En parallèle, ces méthodes peuvent aussi permettre de lire les représentations visuelles chez l'homme, même dans un signal aussi grossier que celui fournit par l'électroencéphalographie, suggérant ainsi l'existence de mécanismes neurophysiologiques communs aux deux espèces.

A l'instar du signal intracrânien enregistré chez le singe, on peut enregistrer des potentiels de champs intracrâniens chez certains patients implantés pour le suivi des crises d'épilepsie. Afin de pouvoir comparer directement les représentations neuronales formées dans le cerveau du singe et celui de l'homme, on fait passer à ces patients la tâche de catégorisation en employant des stimuli identiques à ceux utilisés chez le singe. L'activité présente dans les aires occipito-latérales de l'homme permet un décodage temporel de la catégorie animale similaire à celui observé chez le singe dans la voie visuelle ventrale.

Ensemble, ces démonstrations comportementales et neurophysiologiques convergent vers une origine commune de l'abstraction visuelle entre les singes et l'homme, qui prendrait sa source dans la voie visuelle ventrale des primates. L'existence d'un système neuronal homologue entre ces deux espèces de primates, sous-tendant une fonction de conceptualisation commune, plaide pour une évolution de la cognition et de la continuité de cette fonction cognitive au sein du règne animal jusqu'à l'homme.

4.2 Sur la nécessité d'une approche comparative méticuleuse

Il ne faut pas oublier le terme "comparée " dans "cognition comparée"! Dans le chapitre 2, nous avons essayé de définir ce qu'est la cognition comparée. A travers cet exercice, on se rend compte que la recherche actuelle dans ce domaine consiste plus à faire l'inventaire des différentes capacités cognitives pour chaque espèce, avec un fort biais à chercher des facultés « humaines », plutôt qu'à vraiment comparer directement les modalités cognitives de plusieurs espèces. La comparaison s'effectue à posteriori, dans des méta-analyses, où l'on liste les espèces possédant, ou pas, telle ou telle capacité cognitive, ainsi que le degré de développement de ces capacités. De telles compilations d'études, comme celle présentée en introduction de cette thèse sur la catégorisation, nous renseignent globalement sur la « prévalence » d'une fonction cognitive au sein du règne animal, mais ne permettent en rien d'en retracer l'origine évolutive. En effet, la phylogénie ne peut se déduire que d'homologies, et le fait d'intégrer des études utilisant des protocoles différents, avec des phases d'entraînement aux durées variables, des dispositifs expérimentaux et des stimuli divers, ne laisse guère de chance pour pouvoir trancher entre un mécanisme partagé, hérité d'un ancêtre commun, ou la simple convergence d'adaptations, face à des pressions du milieu similaires. Sans parler des possibles stratégies alternatives que les animaux peuvent mettre en place pour résoudre une tâche et qui n'ont rien à voir avec celles que l'homme utiliserait.

Selon mon point de vue, la seule manière de révéler de vraies homologies ou divergences est la comparaison méticuleuse. En cognition, on étudie un phénomène invisible. On ne peut pas, comme en anatomie, mesurer la taille d'un os ou, comme en génétique, compter le nombre d'acides nucléiques en commun d'un gène de manière indépendante chez plusieurs espèces, pour ensuite mutualiser cette information et en déduire l'arbre phylogénétique des espèces étudiées. La manière la plus rigoureuse, pour comparer cette entité immatérielle, qu'est une fonction cognitive, est de se placer dans des conditions comparables. L'expérience optimale de cognition comparée consiste donc à engager plusieurs espèces dans un protocole et un dispositif expérimental similaire, face aux mêmes stimuli, en utilisant des réponses comportementales aussi similaires que possible (en tenant compte des systèmes perceptifs et moteurs propres à chaque espèce) ainsi qu'un système de renforcement

équivalent. C'est grâce à cette approche que l'on soutient ici, avec conviction, l'utilisation de représentations similaires entre l'homme et le singe pour rapidement extraire une catégorie visuelle d'images naturelles. Sans être une preuve irréfutable, on fournit ainsi un éventail d'arguments forts en faveur de l'existence de mécanismes d'abstractions partagés entre les deux primates. On pose ainsi sérieusement la question d'une homologie de cette fonction cognitive, qui serait par conséquent héritée d'un ancêtre commun.

Similarités cognitives

Nous ne sommes évidemment ni les premiers, ni les seuls, à utiliser cette méthode de comparaison directe entre l'homme et le singe pour rechercher des fonctions homologues. Le traitement des visages est particulièrement important pour la communication et les interactions sociales chez les primates. Effectivement, la comparaison directe de différentes espèces engagées dans un protocole similaire de comparaison visuelle appariée, a permis de montrer qu'hommes et macaques traitent les visages de leur propre espèce de manière spécifique (Pascalis et Bachevalier 1998). Une caractéristique que l'on retrouve chez d'autres primates sociaux comme les macaques de Tonkean et les capucins (Dufour et al. 2006).

Plus spécifiquement, toujours en utilisant des conditions rigoureusement similaires entre les deux espèces, Dahl et al. (2009) démontrent que 3 macaques perçoivent les visages de leurs conspécifiques de la même manière qu'un groupe de 12 humains contrôles. Les deux espèces ont tendance à passer plus de temps à regarder les yeux dans des photographies montrant des visages de primates de la même espèce. Cet effet persiste lorsque que les stimuli sont floutés mais disparaît quand les visages sont présentés à l'envers. En plus de posséder une capacité à traiter spécifiquement les visages conspécifiques, les deux espèces de primates semblent donc avoir hérité d'un système de traitement holistique (global) des visages. Cette similitude dans le traitement des visages est pourtant sujette à controverse (Martin Malivel et al. 2006).

Mais une autre étude renchérit sur la possibilité d'un traitement homologue des visages entre l'homme et le macaque (Adachi et al. 2009). Même si dans cette étude les auteurs ne comparent pas directement les performances des singes à celles de sujets humains, ils utilisent un effet bien connu chez l'homme, l'effet Thatcher, et se

mettent dans des conditions expérimentales les plus comparables possibles avec celle utilisées chez l'homme. Les résultats confirment un traitement configural des visages, similaires entre les deux primates. Il est intéressant de noter qu'une autre étude, chez le babouin, n'avait pas trouvé l'effet Thatcher exposé ici (Parron et Fagot 2008), la différence entre ces deux études étant que Parron et Fagot utilisent un protocole d'appariement différé qui demande un long entraînement pour les singes alors que le protocole d'Adachi en ayant recours au temps de regard, permet de tester l'effet dans des conditions spontanées comme lorsqu'il est observé chez l'homme (Thompson 1980).

Divergences cognitives ?

Un nombre croissant d'études, comparant directement homme et singe dans des protocoles expérimentaux similaires, permet de soutenir l'existence de briques cognitives communes héritées d'un ancêtre commun. Les exemples cités ici argumentent en particulier pour des concepts visuels similaires correspondant à des catégories d'objets socialement et écologiquement pertinentes comme les visages conspécifiques et les animaux. Néanmoins, selon les datations issues des analyses génétiques, l'existence d'un ancêtre commun entre l'homme et les singes de l'ancien monde, que sont les macaques et les babouins, remonte à environ 25-30 millions d'années. Autant de millions d'années pendant lesquelles ces espèces ont évoluées indépendamment. Si la comparaison méticuleuse nous permet de révéler des briques cognitives communes, elle doit aussi permettre de mettre en avant certaines divergences cognitives résultant de cette évolution indépendante.

Nous détaillerons ici deux exemples relatifs aux traitements visuels globaux vs locaux et à la capacité de former des concepts relationnels. En 1977, dans une série d'expériences, Navon démontre qu'il existe en vision un phénomène de précédence du global sur le local. Face à des grandes lettres composées de petites, l'homme est plus rapide pour détecter les grandes lettres et, si on oblige les sujets à catégoriser les petites lettres, ils sont influencés par la congruence avec la lettre globale, alors que l'effet inverse n'est pas observé. Vingt ans plus tard, Fagot et Deruelle (1997) engagent conjointement 14 humains et 8 babouins dans le même type de tâche en utilisant de gros symboles composés de petits symboles. Chez les sujets humains, ils répliquent l'effet de précédence du global sur le local, par contre chez les babouins,

non seulement ils ne répliquent pas cet effet, mais au contraire ils constatent une précédence du local sur le global. L'origine de cette divergence cognitive reste à élucider. Cependant, comme le soulignent les auteurs, ces résultats tendent à démontrer que l'effet de précédence du global sur le local dans la tâche de Navon n'est certainement pas due à une propriété bas-niveau du système visuel mais a de fortes chances d'être influencé par d'autres stratégies cognitives spécifiques de l'homme.

Deux études récentes, dont la nôtre, ouvrent quelques pistes pour comprendre cette divergence. Dans notre étude comportementale, on montre une influence similaire du contexte global sur le traitement local de l'objet entre l'homme et le macaque. En se plaçant dans des conditions plus naturelles pour le système visuel, on observe donc la même précédence du global sur local entre les deux primates. Ceci suggère que la différence qu'observent Fagot et Deruelle (1997) serait liée à la tâche et aux stimuli artificiels qu'ils utilisent. De plus, Davidoff et al. (2008) en testant des individus d'une tribu de Namibie (Himba) montre aussi un avantage du local sur le global dans la tâche de Navon. Ils appuient ainsi l'hypothèse d'une origine culturelle de cet effet plutôt que celle d'un bagage cognitif spécifique à l'être humain.

Dans l'introduction, nous avons parlé des concepts relationnels qui permettent de regrouper des objets non pas par leur similarité physique mais par la nature de la relation qu'ils partagent, telles que l'identité ou la différence. Ce type de concepts n'émerge chez l'enfant que vers l'âge de 4-5 ans et n'a été mis en évidence de manière convaincante que chez les chimpanzés à qui on a appris une forme de langage (Premack 2010).

Toujours en ayant recours à une comparaison directe entre babouins et hommes, Fagot et al. (2001) documentent une des raisons possibles à cette divergence. Les auteurs ont recours à une tâche typique pour tester la capacité à former des concepts relationnels qui est celle de l'appariement différé par relation d'identité ou de différence.

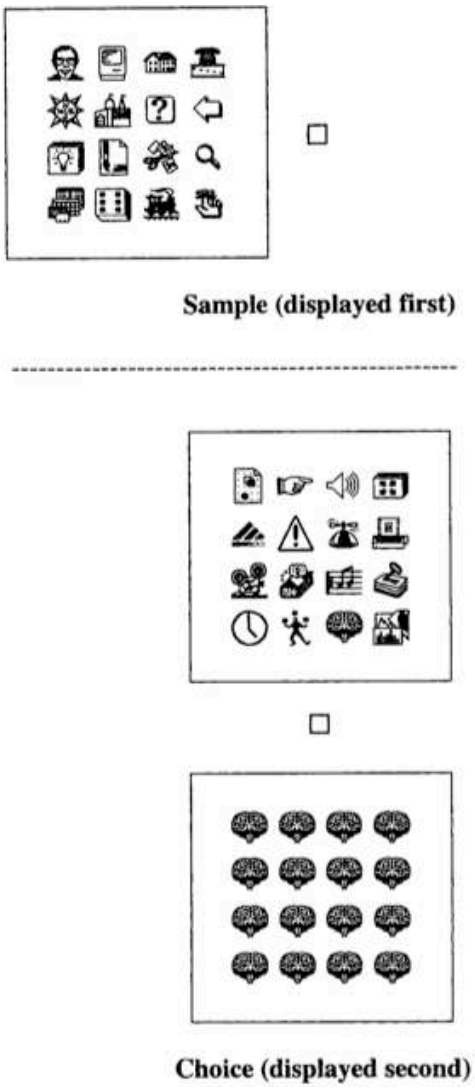

Figure 20: Exemple de matrices de stimulus utilisées dans Fagot et al. (2001). Dans un premier temps une matrice exemple (sample) est présentée. Ensuite deux matrices tests (choice) sont affichées. Le sujet doit choisir celle qui partage une relation d'identité/différence avec la matrice exemple.

Dans ce test, une matrice de 16 stimuli est présentée à 2 babouins et 2 hommes, les stimuli peuvent être soit tous identiques soit totalement ou partiellement différents. Dans une phase de test deux nouvelles matrices, ne contenant aucun stimulus en commun avec la première, sont présentées au sujet qui doit choisir celle qui partage une relation d'identité/différence avec la première matrice présentée. Contrairement à ce qui était supposé à l'époque (Premack 1983, Thompson et Oden 2000), les babouins réussissent cette tâche. Néanmoins une différence notable apparait entre humains et babouins lorsque les auteurs diminuent petit à petit le nombre d'items de la matrice de 16 à 2. Alors que les humains continuent à réussir parfaitement la tâche, les performances des singes tombent au niveau de la chance lorsqu'il s'agit d'apparier un des stimuli en suivant une relation de différence mais réussissent tout de même la

tâche dans le cas d'une relation d'identité ; un résultat que l'on peut prédire en admettant que les singes utilisent l'entropie présente au sein d'une matrice d'item. En effet, qu'il y ait 2 ou 16 items identiques l'entropie est de 0. Par contre, lorsque que seulement deux stimuli sont différents, l'entropie est positive mais beaucoup plus faible en valeur que lorsque il y a 16 stimuli différents, et donc plus difficilement discriminable d'une matrice présentant une entropie de 0. Alors que les auteurs avaient, à l'époque, interprété leurs résultats comme démonstratifs d'une similarité entre les deux primates tout en admettant certains traitements additionnels chez l'homme, Flemming et al. (2007) répliquent ces résultats chez le macaque rhésus et y voient une claire divergence cognitive. Là où les hommes utilisent un concept relationnel abstrait, les singes n'utilisent qu'une simple mesure du désordre présent dans la présentation visuelle. Ces résultats, entre autres, ont amené certains auteurs à prôner une réelle discontinuité cognitive entre l'homme et l'animal, due à une absence chez ce dernier de la capacité à manipuler mentalement des symboles et des relations entre symboles (Penn et al. 2008).

Mais l'histoire n'est peut être pas si simple. D'abord, il est impossible de contrôler l'entropie des stimuli utilisés dans ces tâches d'identité/différence. L'entropie de stimuli symbolisant une différence est toujours supérieure à celle de ceux symbolisant l'identité. Par conséquent, selon le raisonnement de Flemming et al. (2007), toutes les études utilisant ce protocole, même chez l'homme, peuvent aussi être remises en cause. De plus, deux études récentes de l'équipe de Fagot, montrent que certains babouins sont bien capables d'extraire une relation d'identité ou de différence de paires de stimuli (Fagot et Parron 2010, Fagot et Thompson 2011). Ces études révèlent plusieurs facteurs clefs omis par les études précédentes rapportant un résultat négatif. Le facteur le plus important semble être la proximité spatiale entre les deux items dont la relation est à extraire. Effectivement, en partant du constat que les babouins avaient une tendance à se focaliser sur l'information locale (Fagot et Deruelle 1997), les auteurs formulent l'hypothèse selon laquelle l'échec des babouins dans la tâche d'appariement différé relationnel sur des paires d'items, découlerait en fait d'une incapacité à considérer ensemble des stimuli trop éloignés. Et effectivement, en réduisant l'espace entre les 2 items à 2 pixels, les singes sont parfaitement capables de réaliser la tâche. En décortiquant minutieusement les processus cognitifs élémentaires nécessaires pour réussir cette tâche complexe

d'appariement relationnel, Fagot et al. ont pu mettre en évidence le processus à l'origine de la différence entre les hommes et les singes : la faculté de groupement d'objets visuels (Fagot et Parron 2010).

Mais d'autres facteurs entrent en compte dans cette capacité à former des concepts relationnels (Fagot et Thompson 2011). Effectivement, grâce au dispositif expérimental « self-service » décrit dans la partie méthode du chapitre 2, Fagot et Thompson (2011) ont la possibilité de tester de nombreux animaux. Sur les 29 testés, seulement 6 réussissent leur apprentissage dans la tâche d'appariement relationnel et transfèrent à de nouvelles paires de stimuli. La capacité à manipuler des relations symboliques semble donc être fluctuante à travers les différents individus d'un groupe de babouins. Le dispositif expérimental unique utilisé par les auteurs leur permet aussi de faire passer un très grand nombre d'essais au singe. Ainsi pour atteindre une précision de 80% les singes réalisent de 13 000 jusqu'à 30 000 essais. Ce grand nombre d'essai est très laborieux à atteindre avec les méthodes classiques de conditionnement opérant et peut avoir été la source des résultats négatifs rapportés dans certaines autres études. Même si ce nombre d'essais paraît considérable, il est certainement à la mesure du nombre de situations dans lequel un enfant s'est retrouvé à manier ce type de concept, au cours de ses 5 premières années, et qui lui ont permis de développer un raisonnement analogique.

Pour conclure, l'approche comparative méticuleuse est un formidable outil pour mettre en évidence l'existence d'homologies ou de divergences cognitives. Néanmoins, l'interprétation des résultats fournis par cette approche est à prendre avec la plus grande précaution. Effectivement, les différentes études décrites ci-dessus illustrent bien le danger de toutes conclusions hâtives sur l'existence d'homologies ou de divergences. Les principaux écueils sont l'existence de covariables de bas-niveau, l'utilisation de tâches nécessitant d'autres capacités cognitives que celle que l'on veut tester, la variabilité entre individus et le biais à étudier des fonctions cognitives « humaines » pour lesquelles les sujets humains ont eu un entraînement intensif depuis leur naissance alors que les animaux en sont naïfs.

Quand les animaux nous dépassent

Pour illustrer ce dernier point, prenons un exemple montrant un autre intérêt de l'approche comparée directe. Lorsque les résultats d'une étude de cognition comparée entre l'homme et l'animal aboutissent à une divergence, la plupart du temps, elle est en faveur de l'homme qui démontre des facultés supérieures. Une étude récente vient bousculer cette façon de penser et suggère que parfois les animaux peuvent être cognitivement supérieur aux humains. Il s'agit d'une expérience visant à comparer les capacités de mémoire de travail entre l'homme et le chimpanzé (Inoue et Matsuzawa 2007). Les chercheurs ont d'abord entraîné 6 chimpanzés (3 mères et 3 enfants) à classer des séquences de chiffres arabes par ordre croissant. Une séquence de chiffre apparaît aléatoirement sur un écran et les sujets doivent toucher les chiffres par ordre croissant. Tous les chimpanzés réussissent à réaliser cette tâche et transfèrent spontanément à des séquences de chiffres sur lesquelles ils n'ont jamais été entraînés. Afin de tester leur mémoire de travail les chercheurs introduisent un masque qui vient cacher tous les chiffres avec un délai variable (610 ms, 430 ms ou 210 ms après la présentation des stimuli). Pour comparer, 9 étudiants sont entraînés à réaliser la même tâche dans les mêmes conditions. De manière surprenante, les performances d'un des jeunes chimpanzés, appelé Ayumu, dépassent largement celles de tous les humains dans la condition la plus difficile. Plusieurs explications spéculatives ont était avancées pour expliquer cette différence. Les auteurs parlent de mémoires iconique ou eidétisme ; une capacité de mémorisation hors norme qui existe aussi chez certains jeunes enfants. Humphrey (2012), lui, suggère que les singes pourraient être synesthètes et ainsi s'abstraire du masquage par un carré blanc en utilisant une association chiffre/couleur.

Quelle que soit l'explication, cette expérience fournit un des plus beaux exemples d'une supériorité cognitive de l'animal sur l'homme. Encore une fois, l'histoire ne s'arrête pas là. En 2010, Cook et Wilson (2010) ont postulé que cette différence était due à un entraînement prolongé des chimpanzés et qu'après le même entraînement intensif, 2 étudiants ont réussi à égaler les performances d'Ayumu. Cependant, ce dernier ayant aussi continué à s'exercer de son coté, était capable en 2011 d'ordonner parfaitement une suite de 9 chiffres avec un masquage apparaissant au bout de seulement 60 ms, atteignant ainsi une capacité de loin supérieure à tout être humain,

même très entraîné (Matsuzawa 2009, Humphrey 2012). Il semble donc que certains chimpanzés aient juste une mémoire de travail supérieure à la nôtre.

Ce genre de résultats est extrêmement important parce qu'ils ouvrent une porte de sortie à l'anthropocentrisme, cette tendance que l'on a à voir l'homme perché tout en haut de l'arbre de l'évolution. D'un point de vue physique et morphologique personne n'irait clamer que l'homme est le fruit ultime de l'évolution, il n'a pas de fourrure pour se protéger du froid, pas de crocs ni de griffes pour chasser ou se défendre, pas de pieds préhensibles pour grimper dans les arbres etc … C'est un animal comme les autres, sa place est au bout d'une des branches de l'arbre phylogénétique du vivant, au même niveau que toutes les espèces contemporaines. Mais voilà il y a l'intelligence humaine, supérieure à toutes autres, qui en fait une bête à part. Au mieux, on admet que certaines de nos facultés cognitives existent chez d'autres animaux, mais à l'état embryonnaire, sous-développées. Cette conception anthropocentrique du vivant est biaisée. Elle est biaisée par notre subjectivité. Parce que nous vivons sur la terre, nous avons cru que cette planète était le centre de l'univers jusqu'au XVIéme siècle. De la même manière, parce que nous sommes des hommes, et que nous n'avons la capacité d'accéder directement qu'à notre intellect d'homme, on le pense supérieur à ceux du reste du vivant. La physique copernicienne nous a sorti de l'erreur géocentrique, c'est à la cognition comparée de nous sortir de l'erreur anthropomorphique. Selon moi, la méthode la plus radicale pour faire changer les conceptions scientifiques et l'opinion publique sont des expériences de comparaisons cognitives directes, révélant une supériorité cognitive de l'animal sur l'homme, comme celle menée par l'équipe de Matsuzawa. Ces résultats sont aussi destructeurs pour l'anthropomorphisme, que les observations de Copernic l'ont été pour le géocentrisme.

4.3 Une pensée chez l'animal ? Réponse à la problématique générale

Le cadre théorique du cognitivisme nous enseigne que penser, c'est former et manipuler des représentations mentales (Proust 1997). Dans la problématique générale de ce mémoire, nous avons établi comment, les différentes disciplines des sciences cognitives, pourraient permettre de plaider pour l'existence de telles représentations chez l'animal. Ici, nous reprenons l'ensemble des résultats présentés dans ce document thèse, pour les interpréter dans le cadre cette problématique. La

Figure 21, en regroupant des analyses comportementales, neurales et de modélisation de l'Article 3, nous permet d'appuyer et d'illustrer concrètement cette démonstration.

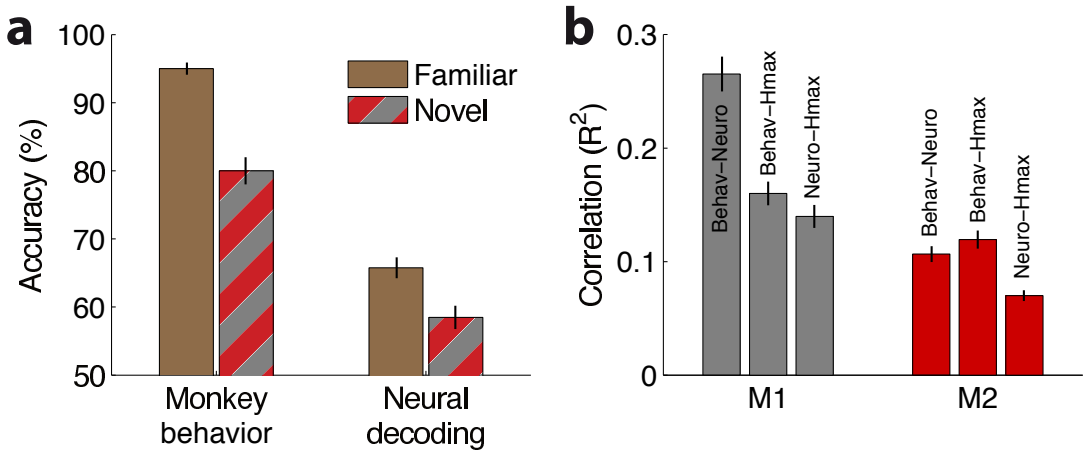

Figure 21: Caractérisation des représentations mentales chez le singe. (a) La capacité de généralisation comportementale et neurale montre l'existence de représentations chez le singe. Les deux singes, tout comme un classifieur entrainé à dissocier le signal neuronal animal non animal évoqué par des images familières, catégorisent des nouveaux stimuli avec des performances bien au dessus de la chance. (b) Ces représentations neurales et comportementales sont liées et sont en partie explicables par un modèle hiérarchique ascendant de la voie visuelle ventrale (Hmax). Pour les deux singes, des corrélations significatives sont observées entre animalité comprotementale et animalité neuronale, entre animalité comportementale et animalité du modèle, ainsi qu'entre animalité neuronale et animalité du modèle.

Grâce à la psychologie expérimentale, nous avons tout d'abord montré que les singes utilisaient des représentations mentales **(Article 1)**. Effectivement, ils sont capables de catégoriser des images comme contenant un animal ou non, quels que soient le type d'animal, sa taille, son excentricité et l'information contextuelle qui l'entoure. Plus important, ils sont capables de réaliser cette tâche sur des images qu'ils n'ont jamais vues auparavant **(Article 1 et 3,** Figure 21 a). On ne peut donc pas expliquer leur comportement par de simples associations stimulus-réponse enregistrées en mémoire. La seule explication cognitive possible est qu'ils possèdent bien une représentation mentale flexible correspondant à la catégorie animale.

Les neurosciences, et en particulier, les méthodes de décodage pour lire dans les pensées, nous ont ensuite permis de caractériser ces représentations au niveau neuronal **(Article 2 et 3)**. La voie ventrale fournit un signal catégoriel précoce qui permet de prédire si le singe a vu ou non un animal, même si c'est la première fois qu'il le voit (Figure 21 a). Ce signal est en lien direct avec les temps de réaction du singe (Figure 21 b). Ces résultats confirment l'existence biologique de représentations

mentales dans le cerveau du singe, auxquelles il a recourt pour résoudre la tâche de catégorisation.
Enfin, l'intelligence artificielle nous fournit des outils pour modéliser ces représentations **(Article 3)**. Un modèle hiérarchique ascendant de la voie visuelle ventrale (Hmax, Serre et al. 2007) permet d'expliquer en partie les performances comportementales et neurales des singes (Figure 21 b).

A l'aide des trois outils du cogniticien que sont la *psychologie expérimentale*, les *neurosciences* et *l'intelligence artificielle,* on a pu *caractériser*, montrer *l'existence biologique* et *modéliser* des représentations mentales chez le singe. Cette démonstration tend à accorder une forme de pensée au singe. Mieux, dans cette thèse nous avons aussi montré, par le comportement, que les hommes et les singes semblent partager ces représentations mentales et qu'elles semblent vraisemblablement codées de la même manière, dans des territoires corticaux homologues **(Article 4 et résultats préliminaires chapitre 3.7)**. Les performances des hommes sont aussi parfaitement modélisables par le modèle Hmax (Serre et al. 2007). Par conséquent, le système représentationnel impliqué dans la tâche de catégorisation rapide apparaît similaire entre l'homme et le singe, à tous les niveaux d'études. On peut donc conclure, que non seulement les singes pensent, mais qu'en plus cette pensée est en partie la même que la nôtre.

4.4 Ouverture et Perspectives

Une conscience pour l'animal ?

Dès le début de cette thèse, nous avons sciemment écarté la notion de conscience de nos travaux et questionnements sur la pensée animale, en se concentrant sur la notion de représentions, plus facilement abordable et opérationnelle à étudier dans le cadre des sciences cognitives. Néanmoins, aussi délicat que le sujet puisse paraître, l'étude de la conscience chez l'animal n'est pas hors de portée d'une approche expérimentale.

La question de la conscience animale est d'ailleurs une des quêtes principales de l'Ethologie Cognitive de Donald Griffin. Pour lui, un animal peut être considéré comme éprouvant un niveau de conscience rudimentaire, s'il pense de manière subjective à un objet ou un événement (Griffin 2001). Bien avant Griffin, à l'heure du

behaviourisme, Tolman (1927) avait déjà proposé d'interpréter les comportements d'hésitations (va et vient du regard et/ou déplacements), qu'un animal peut présenter avant de répondre dans un protocole de conditionnement opérant, comme des marqueurs d'une forme de conscience. Nous avons aussi parlé à plusieurs reprises de la conscience de soi mise en évidence chez les grands singes, certains cétacés, l'éléphant d'Asie ou encore la pie bavarde, grâce au test du miroir de Gallup (1968).

Plus récemment, l'étude de la métacognition (capacité à penser à ses propres pensées), à travers les outils de la psychologie comparée, offre des démonstrations convaincantes et rigoureuses d'une forme de réflexion consciente chez l'animal (voir Smith 2009 pour une revue). Dans la tâche classique de métacognition, l'animal comme l'homme, semble capable de juger de la qualité de ses représentations mentales (perceptuelle, mémorielles), en choisissant de ne pas répondre sur des essais difficiles où il aurait de grandes chances de se tromper. Des hommes engagés dans un protocole similaire montre la même distribution de ces « réponses d'incertitude » et rapportent verbalement que, lors de ces réponses, ils réalisent un acte métacognitif autoréflexif (Smith et al 1997).

Enfin, les avancées majeures dans l'étude des corrélats neuronaux de la conscience chez l'homme (Tononi et Koch 2008) fournissent autant de marqueurs permettant d'évaluer l'existence biologique d'une conscience chez l'animal (Edelman et Seth 2009). Par exemple, la tâche de rivalité binoculaire, qui permet de comparer les substrats neurobiologiques de la perception consciente à ceux de la perception inconsciente des mêmes informations sensorielles, a été utilisée à la fois chez l'homme et le singe. Le phénomène de rivalité binoculaire s'observe lorsque l'on présente à chaque œil un stimulus différent. Or, dans une telle situation, nous ne percevons pas le mélange des deux images, mais alternativement l'une ou l'autre à des intervalles de quelques secondes. Ainsi, la perception de l'un des stimuli est tantôt consciente, tantôt inconsciente, alors que l'entrée visuelle sur la rétine reste la même. Chez le singe, l'activité des neurones d'IT corrèle avec le percept rapporté en comportement par le sujet, tandis que l'activité des neurones de V1 code le signal visuel indépendamment du percept de l'animal (Logothetis 1998). En engageant des patients épileptiques dans une tâche similaire à la rivalité binoculaire, appelée « flash supression », Kreiman et al. (2002) enregistrent l'activité des neurones dans le lobe

temporal médian. Comme pour les neurones d'IT chez le singe, l'activité des deux tiers de ces neurones corrèle avec le percept conscient du patient. Ces résultats, chez l'homme et le singe convergent, pour supporter l'hypothèse selon laquelle, la conscience visuelle émergerait relativement haut dans la hiérarchie visuelle des primates. Ces conclusions sont bien sûr discutables. En particulier, dans l'article 5, nous soulignons l'importance des processus réentrant vers le cortex visuelle primaire, dans la métacognition visuelle. Néanmoins, la comparaison des corrélats neuronales inter-espèces provenant de ces deux études, suggère une continuité biologique de la conscience visuelle. Il reste cependant beaucoup à faire, aussi bien pour comprendre l'émergence de la conscience humaine à partir du substrat biologique, que dans la comparaison intégrative des phénomènes de « conscience » au sein du règne animal.

Une continuité émotionnelle ?

La question de continuité cognitive, à laquelle nous nous sommes intéressés dans cette thèse, fait échos à celle de continuité affective. Est-ce que les animaux ressentent des émotions ? Est-ce que ces émotions sont semblables aux nôtres ?

Bien que, comme la conscience, ces questions soient plus difficiles à objectiver, certains chercheurs s'y sont tout de même « attaqués » (Pankskeep 2004, Bekoff 2000). Les animaux démontrent clairement, à tout observateur attentif, des signes comportementaux qui témoignent de la richesse des émotions qu'ils peuvent ressentir (un chien qui joue, un chat vexé, un cheval qui a peur …). Cependant, on s'est longtemps refusé à leur accorder des émotions, même aussi basiques que la souffrance ou la peur, sûrement pour pouvoir continuer à les manipuler et à les exploiter sans états d'âmes. De nos jours l'étude scientifique du comportement apporte des évidences indéniables de la capacité des animaux à ressentir la peur, la joie, la honte, l'embarras, la jalousie, la rage, la colère, l'amour, le respect, le dégoût ou la tristesse (Bekoff, 2000).

A ces observations comportementales vient s'ajouter le constat neuroscientifique qu'une des structures les mieux conservées entre les différents mammifères est certainement le système limbique, qui contient entre autre l'amygdale, une structure clé pour le traitement des émotions chez l'homme. Etant donné que la plupart des animaux expriment des signes émotionnels évidents et possèdent une structure, qui, chez l'homme, serait responsable du ressenti émotionnel, il serait absurde de se refuser à attribuer aux mammifères non humains des émotions tout aussi fortes que les

nôtres. Et même peut être encore plus fortes, puisque leur ratio entre système limbique et néocortex est supérieur au nôtre (Pankskeep 2011).

Des personnalités animales

Dans l'expérience où l'on présente des images de serpents aux singes, les deux animaux montrent une inhibition pour répondre sur ces images au cours de la première session. Ensuite, l'un des deux s'y habitue et répond normalement sur ces images comme sur n'importe quelle autre image animale cible. En revanche, le deuxième singe n'aura de cesse d'éviter de répondre sur les serpents pendant les 13 séances de l'expérience.

Cette variabilité interindividuelle suggère l'existence de différents traits de personnalité, ou de vie, menant à des performances complètement différentes dans notre protocole de conditionnement opérant. Alors que le premier singe présente une inhibition rapidement effacée par le protocole de renforcement, le deuxième, de part son inhibition irrépressible, présente un comportement assimilable à une phobie des serpents. Cette variabilité des réponses d'inhibition ou de peur chez le macaque face aux serpents a déjà été soulignée par d'autres. Nelson et al. (2003) ont testé 21 singes dans un appareil qui oblige l'animal à attraper sa nourriture préférée sur une boîte en plexiglas. Cette boîte peut contenir un serpent vivant, un faux serpent, un rouleau de scotch ou rien du tout. La plupart des singes mettent plus de temps pour saisir la nourriture quand un serpent est présent dans la boîte. Cependant, cette inhibition peut varier d'un facteur 5 entre les individus. De plus, certains seulement montrent un comportement de rétraction (ils retirent totalement la main de l'appareil) témoignant selon les auteurs d'une réelle peur pour l'animal.

Même si elles ne sont pratiquement jamais prisent en compte dans les protocoles de cognition comparée, il peut exister une grande variabilité dans les réponses de différents individus de la même espèce. L'écologie comportementale, elle, fait une large place à cette variabilité intra-spécifique, qui est l'un des prérequis indispensable à l'évolution des espèces : sans variabilité entre les individus, la sélection naturelle ne pourrait pas sélectionner certains traits mieux adaptés, à l'origine de l'émergence de nouvelles espèces. Ainsi les écologistes comportementalistes n'hésitent pas à parler de personnalité chez l'animal. Dans la très sérieuse revue *Trends in Ecology and Evolution*, Biro et Stamps (2008) nous apprennent que les truites peuvent être plus ou

moins agressives, les cochons audacieux ou timides et les mésanges calmes ou actives. Il est évident que ces différentes personnalités, qu'elles soient génétiques ou liées à l'histoire de vie, peuvent avoir une influence majeure sur les résultats de n'importe quelle étude comportementale. Ces différences entre individus sont à prendre en compte, y compris dans le domaine de la cognition comparée.

Lire les pensées intimes des animaux

On a vu que chez l'homme, les techniques de lecture de pensée, permettaient de lire plusieurs types de représentations. On a notamment opposé les représentations induites par l'environnement, comme les représentations visuelles, aux représentations générées intérieurement, comme les souvenirs ou l'imagerie mentale.

Même si dans une définition cognitiviste, tout acte de traitement de l'information fait partie de ce que l'on appelle la pensée, dans l'opinion publique, la pensée se réfère plus à une activité endogène, abstraite des stimulations environnementales et donc basée sur des représentations générées intérieurement. Or, il existe peu de preuves tangibles ou directes prouvant la capacité des animaux à former ce genre de représentations. Notamment, la capacité d'imagerie mentale indépendante de toutes stimulations externes, chez l'animal, reste une question sans réponse. Les méthodes développées ici pourraient peut être permettre d'apporter de nouveaux éléments sur ce point. Effectivement, nous avons vu chez l'homme que les représentations induites par l'environnement et générées intérieurement peuvent partager des réseaux communs. Par exemple les représentations mises en jeux dans la catégorisation d'objets visuels sont aussi celles qui sont réactivées pendant une tâche d'imagerie mentale (Reddy et al. 2009) ou de rappel spontané (Polyn et al. 2005).

Dans cette thèse, nous avons déjà montré qu'il était possible de lire les représentations catégorielles visuelles chez l'homme et le singe dans les potentiels de champs intracrâniens. Dans une étude que nous venons juste de commencer chez le patient épileptique, en engageant alternativement le patient dans des tâches de catégorisation visuelle et d'imagerie mentale des mêmes objets, nous essayons de montrer une activité spatio-temporelle commune à ces deux types de représentations. Comme Reddy et al. (2009) l'ont montré en IRM fonctionelle, on s'attend, en utilisant le bon signal au bon moment, à pouvoir prédire la catégorie imaginée par le patient, en utilisant uniquement l'activité induite par la présentation visuelle du stimulus. Si ces résultats sont concluants, c'est à dire si chez l'homme l'utilisation de l'activité

évoquée par des stimuli visuels nous permet de détecter l'imagination interne de ces stimuli alors on peut tout aussi bien imaginer décoder les pensées endogènes chez le singe.

Un protocole que l'on pourrait utiliser serait le suivant : on apprendrait au singe à associer différents symboles avec des stimuli importants pour lui (nourriture, congénère, expérimentateur …). Dans une première phase, on présenterait le symbole puis le stimulus pertinent à l'animal. On apprendrait alors à un classifieur à décoder les représentations induites par la présentation visuelle de ces différents stimuli. Dans une deuxième phase, on ne présenterait que les symboles qui seraient suivis d'un écran noir. Si, à l'aide du classifieur entrainé dans la tâche perceptuelle, on est capable de décoder les stimuli associés aux symboles, on mettrait alors en évidence l'existence *d'imagerie mentale induite* chez le singe. Enfin, dans une dernière phase, on enregistrerait l'activité spontanée du singe pendant qu'il fixerait un écran noir durant plusieurs minutes. En testant, l'algorithme sur toute cette période de repos, on pourrait ainsi savoir si le singe réactive spontanément les représentations de nourriture ou de congénères. On aurait alors mis en évidence que pendant qu'il fixe cet écran noir, le singe fait de *l'imagerie mentale spontanée ;* ou plus simplement, qu'il est en train de penser à autre chose.

Dans une logique similaire, on peut imaginer un autre protocole utilisant le principe du neurofeedback. Des résultats récents chez l'homme (Cerf et al. 2010) montrent qu'en utilisant l'activité de neurones situés dans le lobe temporal médian, des patients sont capables de faire apparaître petit à petit l'image à laquelle ils pensent.
Dans ce deuxième protocole fictif, on conserverait la première phase du protocole précédent. Dans la deuxième phase, au lieu de laisser l'écran noir après la présentation du symbole, on présenterait le stimulus mais légèrement bruité (rajout de bruit blanc ou randomisation de phase). Grâce à un classifieur en temps réel, on ne récompenserait le singe que s'il active un pattern neural approchant au maximum celui de la présentation du stimulus visuel non bruité. Si cela fonctionne, on pourrait, petit à petit, augmenter le niveau de bruit, jusqu'à faire disparaître totalement l'image. A la fin de l'expérience, même en présence d'un stimulus visuel constitué à 100% de bruit, le singe réactiverait la représentation de l'objet. De tels résultats pourraient être interprétés comme une capacité *d'imagerie mentale conditionnée.*

5 Annexe

5.1 « The only real valuable thing is intuition ». Albert Einstein

J'ai écrit cette annexe dans un seul but : Si jamais par le plus grand des hasards un étudiant ouvrait ce document de thèse et si, par un hasard encore plus grand, il arrivait jusqu'à cette annexe, le seul message que j'aimerais qu'il retienne serait le suivant : *il faut croire en ses intuitions*! Pas parce qu'elles sont toujours bonnes, mais parce qu'elles constituent une source de motivation inégalable pour pousser une idée embryonnaire jusqu'à la potentielle publication d'un résultat scientifique, à travers le long et douloureux chemin de l'expérimentation, de l'analyse et de l'écriture.

Petite Anecdote

Il y a bien longtemps, les étudiants de Master 1 en neuroscience et comportement animal de la faculté des sciences de Toulouse, avaient la chance incroyable de monter eux-mêmes un projet de recherche d'une durée d'environ trois mois, visant à valider leurs capacités pratiques et expérimentales.

Après 21 ans d'encadrement rapproché dans des disciplines aussi classiques qu'ennuyeuses, le système éducatif nous donnait enfin un espace de liberté personnel dans ma discipline chérie: la cognition animale. Il ne me fallut pas longtemps pour prendre contact avec le zoo le plus proche et les supplier de nous laisser étudier leurs chimpanzés. A ma grande surprise, ils acceptèrent. Nous nous démenâmes alors pour fabriquer un dispositif expérimental à la mesure de la force de ces grands singes, à grand coup de perceuse et de scie sauteuse. Pendant toute la mise au point du protocole, nous eûmes la chance d'observer ces animaux magnifiques au regard aussi profond que triste. Et hasard ou coïncidence, je pus à nouveau constater à quel point ils étaient adroits lorsqu'il s'agissait d'atteindre une cible avec un projectile. Cette fois, c'est un vieux mâle qui, après trois-quatre bruyants allers-retours, suspendu au sommet de sa cage ridiculement petite, glissa sa main à travers les barreaux et, par un déroulement habile de l'avant bras, projeta un excrément dans la direction de mon visage. Suffisamment loin de l'enclos, j'eus le temps de voir arriver le projectile et de l'éviter au dernier moment, dans un mouvement de tête matricien. La trainée marron sur mon col de chemise témoignait du danger que je venais d'éviter. Malheureusement, ou heureusement peut être, au dernier moment, sans nous

expliquer vraiment pourquoi, la direction du zoo nous fit comprendre qu'elle revenait sur sa décision et qu'on était prié de faire nos expériences ailleurs. C'est comme ça que je me suis retrouvé au CerCo. Grâce à Franck Lefèvre, alors en charge de l'animalerie, et à Michèle, nous pûmes transférer notre protocole d'étude de l'empan mnésique aux macaques rhésus du laboratoire.

Au cours de ce premier stage que je découvrais une des thématiques phare de l'unité de recherche, la catégorisation rapide. Et pas n'importe laquelle, celle des animaux. Autrement dit, celle que je pratiquais depuis tout petit au cours des nombreuses sorties naturalistes que j'avais effectué avec mon père. Des moments pendant lesquels, je m'étais souvent surpris à détecter, presque inconsciemment, la présence d'un chevreuil ou d'un renard à plusieurs centaines de mètres, malgré la faible lumière des affuts crépusculaires.

Mais ce qui m'avait le plus marqué, et à plusieurs reprises, c'était ma capacité et celle de mon chien Nouck, à percevoir quasiment instantanément l'ondulation d'un serpent dans l'herbe. Je n'ai pas peur des serpents, au contraire, si je suis assez rapide j'aime bien les attraper pour les toucher et les regarder. Cependant, ma première réaction, quand j'ai la chance d'en apercevoir un, est systématiquement une réaction d'évitement. J'ai comme un mouvement de recul et je sens mon cœur s'accélérer. Cette réaction ne doit durer qu'une fraction de seconde puisque bien souvent, je me jette aussitôt dans l'herbe pour essayer de me saisir de l'animal. Il m'est arrivé plus d'une fois de ressentir cette même sensation de peur fulgurante, en face d'un tuyau d'arrosage ou d'une branche aux formes sinueuses.

J'étais sûr de mon coup, s'il y a bien quelque chose qu'on reconnaît rapidement ce sont les serpents. Inconscient du danger, je décidais de parler de cette hypothèse à Michèle, la directrice du laboratoire, et j'eus la bonne idée d'ajouter, comme argument pour la convaincre, que les singes possédaient une peur innée des serpents, information que j'avais lue quelque part ou bien que mon père m'avait racontée. J'appris deux choses de cette discussion : qu'il ne fallait pas déranger Michèle sans la prévenir à l'avance et que les scientifiques, contrairement aux journalistes, devaient TOUJOURS citer leurs sources. Qui plus est, il est conseillé que ces sources soient des articles scientifiques reconnus.

Ce n'est que récemment, en me replongeant dans cette thématique, que j'ai effectivement découvert une série d'articles rapportant que des singes élevés en laboratoire n'ont pas une peur innée, *stricto sensu*, des serpents. Par contre, ils présentent des prédispositions à acquérir cette peur par l'observation du comportement de leurs congénères. Si on leur montre une vidéo d'un singe effrayé par un serpent, ils développent instantanément une frayeur intense pour les serpents alors que la même vidéo, dans laquelle le serpent est remplacé par une fleur, n'entraîne aucune réaction de peur face aux fleurs (Cook et Mineka 1989).

J'étais donc très refroidi sur mon hypothèse du traitement rapide des serpents, et les nombreuses expériences, que j'imaginais déjà, tombaient à l'eau. Je finis mon stage de master avec des résultats mitigés sur l'empan mnésique des rhésus, puis, embrayait avec le master 2 et la thèse, en m'appliquant à faire passer rigoureusement les expériences préalablement définies avec Denis.

Mais, pendant l'été de ma première année de thèse, j'eus l'occasion d'établir et de mener une expérience tout seul. Je profitais de la désertification aoûtiennes du laboratoire pour glisser, en douce, quelques images de serpents dans les stimuli de cette expérience. Après quelques sessions d'enregistrement, je m'empressais d'analyser les résultats. A ma grande joie, je découvris les effets comportementaux ainsi que les différences en potentiels évoquées reportés dans cette thèse. Extrêmement enthousiasmé, j'envoyais par email le fruit de mon labeur à Denis, Thomas et Gabriel, afin de leur faire partager cette joie euphorique. J'eus alors droit à ma deuxième douche froide. On souligna l'inintérêt de cette question et le manque de contrôles rigoureux.

Et pourtant, comme vous avez pu vous en apercevoir, les résultats de cette incartade sont présentés à plusieurs endroits de ce manuscrit et constituent des points de discussion clefs. Pire, grâce à Michèle un article est en cours. Merci Michèle !

BIBLOGRAPHIE

Adachi, I., Chou, D. P., & Hampton, R. R. (2009). Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Current biology : CB*, *19*(15), 1270–3. doi:10.1016/j.cub.2009.05.067

Afraz, S.-R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, *442*(7103), 692–5. doi:10.1038/nature04982

Anderson, J. R., & Gallup, G. G. (2011). Do rhesus monkeys recognize themselves in mirrors? *American journal of primatology*, *73*(7), 603–6. doi:10.1002/ajp.20950

Anderson, J. R., Myowa-Yamakoshi, M., & Matsuzawa, T. (2004). Contagious yawning in chimpanzees. *Proceedings. Biological sciences / The Royal Society*, *271 Suppl* , S468–70. doi:10.1098/rsbl.2004.0224

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *The Behavioral and brain sciences*, *33*(4), 245–66; discussion 266–313. doi:10.1017/S0140525X10000853

Arcizet, F., Jouffrais, C., & Girard, P. (2008). Natural textures classification in area V4 of the macaque monkey. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *189*(1), 109–20. doi:10.1007/s00221-008-1406-9

Ashby, F., & Lee, W. (1993). Perceptual variability as a fundamental axiom of perceptual science. *Advances in Psychology*

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature neuroscience*, *7*(5), 542–8. doi:10.1038/nn1241

Aust, U., & Huber, L. (2002). Target-defining features in a “people-present/people-absent” discrimination task by pigeons. *Animal learning & behavior*, *30*(2), 165–76.

Avarguès-Weber, A, Portelli, G., Benard, J., Dyer, A., & Giurfa, M. (2010). Configural processing enables discrimination and categorization of face-like stimuli in honeybees. *The Journal of experimental biology*, *213*(4), 593–601. doi:10.1242/jeb.039263

Avarguès-Weber, Aurore, Dyer, A. G., Combe, M., & Giurfa, M. (2012). Simultaneous mastering of two abstract concepts by the miniature brain of bees. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7481–6. doi:10.1073/pnas.1202576109

Avarguès-Weber, Aurore, Dyer, A. G., & Giurfa, M. (2011). Conceptualization of above and below relationships by an insect. *Proceedings. Biological sciences / The Royal Society*, *278*(1707), 898–905. doi:10.1098/rspb.2010.1891

Awazu, S., & Fujita, K. (2006). Transfer from “edible” categorization training to feeding behavior in pigeons (Columba livia) 1. *Japanese Psychological Research*.

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), 1013–26. doi:10.1037/0096-1523.33.5.1013

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–54. doi:10.1073/pnas.0507062103

Barbeau, E. J., Taylor, M. J., Regis, J., Marquis, P., Chauvel, P., & Liégeois-Chauvel, C. (2008). Spatio temporal dynamics of face recognition. *Cerebral cortex (New York, N.Y. : 1991)*, *18*(5), 997–1009. doi:10.1093/cercor/bhm140

Bates, L. A., Sayialel, K. N., Njiraini, N. W., Moss, C. J., Poole, J. H., & Byrne, R. W. (2007). Elephants classify human ethnic groups by odor and garment color. *Current biology : CB*, *17*(22), 1938–42. doi:10.1016/j.cub.2007.09.060

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*.

Bekoff, M. (2001). Observations of scent-marking and discriminating self from others by a domestic dog (Canis familiaris): tales of displaced yellow snow. *Behavioural processes*, *55*(2), 75–79.

Benard, J., Stach, S., & Giurfa, M. (2006). Categorization of visual stimuli in the honeybee Apis mellifera. *Animal cognition*, *9*(4), 257–70. doi:10.1007/s10071-006-0032-9

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

Bentin, S., Taylor, M. J., Rousselet, G. A., Itier, R. J., Caldara, R., Schyns, P. G., Jacques, C., et al. (2007). Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nature neuroscience*, *10*(7), 801–2; author reply 802–3.

Bhatt, R., Wasserman, E., Reynolds, W., & Knauss, K. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: animal behaviour processes*.

Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *The Journal of comparative neurology*, *306*(4), 554–75. doi:10.1002/cne.903060403

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science (New York, N.Y.)*, *282*(5389), 746–9.

Bruner, J., Goodnow, J., & Austin, G. (1956). A study of thinking.

Bullier, J. (2001). Integrated model of visual processing. *Brain research. Brain research reviews*, *36*(2-3), 96–107.

Cook,P.,Willson, J. (2011). Do young chimpanzees have extraordinary working memory? *Psychonomic bulletin & review*, *17*(4), 599–600.

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). doi:10.1167/11.10.9

Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R. Q., Koch, C., & Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature*, *467*(7319), 1104–8. doi:10.1038/nature09510

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current biology : CB*, *20*(6), 544–7. doi:10.1016/j.cub.2010.01.053

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), 913–9. doi:10.1038/13217

Chun, M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, affective & behavioral neuroscience*, *4*(2), 117–26.

Collins, C. E., Airey, D. C., Young, N. A., Leitch, D. B., & Kaas, J. H. (2010). Neuron densities vary across and within cortical areas in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(36), 15927–32. doi:10.1073/pnas.1010356107

Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of abnormal psychology*, *98*(4), 448–59.

Cook, P., & Wilson, M. (2010). Do young chimpanzees have extraordinary working memory? *Psychonomic bulletin & review*, *17*(4), 599–600. doi:10.3758/PBR.17.4.599

Cook, R., & Fagot, J. (2009). First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9530–3. doi:10.1073/pnas.0903378106

Cook, R. G., Katz, J. S., & Blaisdell, A. P. (2012). Temporal properties of visual search in pigeon target localization. *Journal of experimental psychology. Animal behavior processes*, *38*(2), 209–16. doi:10.1037/a0026496

Coulon, M., Deputte, B. L., Heyman, Y., & Baudoin, C. (2009). Individual recognition in domestic cattle (Bos taurus): evidence from 2D-images of heads from different breeds. *PloS one*, *4*(2), e4441. doi:10.1371/journal.pone.0004441

Coulon, M., Deputte, B. L., Heyman, Y., Delatouce, L., Richard, C., & Baudoin, C. (2007). Visual discrimination by heifers (Bos taurus) of their own species. *Journal of comparative psychology (Washington, D.C. : 1983)*, *121*(2), 198–204. doi:10.1037/0735-7036.121.2.198

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

Dahl, C. D., Wallraven, C., Bülthoff, H. H., & Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Current biology : CB*, *19*(6), 509–13. doi:10.1016/j.cub.2009.01.061

David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of neurophysiology*, *96*(6), 3492–505. doi:10.1152/jn.00575.2006

Davidoff, J., Fonteneau, E., & Fagot, J. (2008). Local and global processing: Observations from a remote culture. *Cognition*, *108*, 702–709. doi:10.1016/j.cognition.2008.06.004

Delfour, F., & Marten, K. (2001). Mirror image processing in three marine mammal species: killer whales (Orcinus orca), false killer whales (Pseudorca crassidens) and California sea lions (Zalophus californianus). *Behavioural processes*, *53*(3), 181–190.

Deloache, J. S., & Lobue, V. (2009). The narrow fellow in the grass: human infants associate snakes and fear. *Developmental science*, *12*(1), 201–7. doi:10.1111/j.1467-7687.2008.00753.x

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), 2187–200.

Delorme, Arnaud, & 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. doi:10.1016/j.jneumeth.2003.10.009

Delorme, Arnaud, Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in psychology*, *1*(June), 21. doi:10.3389/fpsyg.2010.00021

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), 2051–62.

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

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), 2470–3. doi:10.1126/science.1063414

Dufour, V., Pascalis, O., & Petit, O. (2006). Face processing limitation to own species in primates: a comparative study in brown capuchins, Tonkean macaques and humans. *Behavioural processes*, *73*(1), 107–13. doi:10.1016/j.beproc.2006.04.006

Dufour, V., Wascher, C. a F., Braun, A., Miller, R., & Bugnyar, T. (2012). Corvids can decide if a future exchange is worth waiting for. *Biology letters*, *8*(2), 201–4. doi:10.1098/rsbl.2011.0726

D'amato, M., & Van Sant, P. (1988). The person concept in monkeys (Cebus apella). *Journal of Experimental Psychology: animal behaviour processes*.

Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends in neurosciences*, *32*(9), 476–84. doi:10.1016/j.tins.2009.05.008

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. doi:10.1038/33402

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current biology : CB*, *19*(12), 1028–33. doi:10.1016/j.cub.2009.04.054

Evarts, E. (1968). A technique for recording activity of subcortical neurons in moving animals. *Electroencephalography and clinical neurophysiology*.

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), 171–80.

Fabre-Thorpe, M, Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, *9*(2), 303–8.

Fabre-Thorpe, Michèle. (2003). Visual categorization: accessing abstraction in non-human primates. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1435), 1215–23. doi:10.1098/rstb.2003.1310

Fabre-Thorpe, Michèle. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in psychology*, *2*(October), 243. doi:10.3389/fpsyg.2011.00243

Fagot, J, & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (Homo sapiens) and baboons (Papio papio). *Journal of experimental psychology. Human perception and performance*, *23*(2), 429–42.

Fagot, J, Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (Papio papio) and humans (Homo sapiens). *Journal of experimental psychology. Animal behavior processes*, *27*(4), 316–28.

Fagot, Joël, & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior research methods*, *42*(2), 507–16. doi:10.3758/BRM.42.2.507

Fagot, Joël, & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(46), 17564–7. doi:10.1073/pnas.0605184103

Fagot, Joël, & Parron, C. (2010). Relational matching in baboons (Papio papio) with reduced grouping requirements. *Journal of experimental psychology. Animal behavior processes*, *36*(2), 184–93. doi:10.1037/a0017169

Fagot, Joël, & Thompson, R. K. R. (2011). Generalized Relational Matching by Guinea Baboons (Papio papio) in Two-by-Two-Item Analogy Problems. *Psychological science*. doi:10.1177/0956797611422916

Fagot, Joël, Thompson, R. K. R., & Parron, C. (2010). How to read a picture: lessons from nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 519–20. doi:10.1073/pnas.0913577107

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, *9*, 1871–1874.

Farah, M. J. (1992). Agnosia. *Current opinion in neurobiology*, *2*(2), 162–4.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, *70*(6), 510–23.

Felleman, D. J., & Van Essen, D. C. (n.d.). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *1*(1), 1–47.

Ferreira, G., Keller, M., Saint-Dizier, H., Perrin, G., & Lévy, F. (2004). Transfer between views of conspecific faces at different ages or in different orientations by sheep. *Behavioural processes*, *67*(3), 491–9. doi:10.1016/j.beproc.2004.08.005

Fiorito, G., von Planta, C., & Scotto, P. (1990). Problem solving ability of Octopus vulgaris Lamarck (Mollusca, Cephalopoda). *Behavioral and neural biology*, *53*(2), 217–30.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499–504.

Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (Macaca mulatta): judgment of relations and relations-between-relations. *Journal of experimental psychology. Animal behavior processes*, *33*(1), 55–63. doi:10.1037/0097-7403.33.1.55

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), 312–6. doi:10.1126/science.291.5502.312

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), 845–51. doi:10.1126/science.1194908

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science (New York, N.Y.)*, *259*(5091), 100–3.

Gardner, R., & Gardner, B. (1969). Teaching sign language to a chimpanzee. *Science*.

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.

Geach, P. (1957). Mental acts; their content and their objects.

Ghosh, N., Lea, S. E. G., & Noury, M. (2004). Transfer to intermediate forms following concept discrimination by pigeons: chimeras and morphs. *Journal of the experimental analysis of behavior*, *82*(2), 125–41. doi:10.1901/jeab.2004.82-125

Girard, P, Jouffrais, C., & Kirchner, C. H. (2008). Ultra-rapid categorisation in non-human primates. *Animal cognition*, *11*(3), 485–93. doi:10.1007/s10071-008-0139-2

Girard, Pascal, & Koenig-Robert, R. (2011). Ultra-rapid categorization of fourier-spectrum equalized natural images: macaques and humans perform similarly. *PloS one*, *6*(2), e16453. doi:10.1371/journal.pone.0016453

Giurfa, M., Eichmann, B., & Menzel, R. (1996). Symmetry perception in an insect. *Nature*, *382*(6590), 458–61.

Giurfa, M., Vorobyev, M., Brandt, R., Posner, B., & Menzel, R. (1997). Discrimination of coloured stimuli by honeybees: alternative use of achromatic and chromatic signals. *Journal Of Comparative Physiology*.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature*, *410*(6831), 930–3. doi:10.1038/35073582

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (Papio papio). *Science (New York, N.Y.)*, *336*(6078), 245–8. doi:10.1126/science.1218152

Griffin, D. (2001). Animal minds: Beyond cognition to consciousness.

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.

Hampton, A. N., & O'doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1377–82. doi:10.1073/pnas.0606297104

Harley, H. E., Roitblat, H. L., & Nachtigall, P. E. (1996). Object representation in the bottlenose dolphin (Tursiops truncatus): integration of visual and echoic information. *Journal of experimental psychology. Animal behavior processes*, *22*(2), 164–74.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–5. doi:10.1038/nature07832

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 (New York, N.Y.)*, *293*(5539), 2425–30. doi:10.1126/science.1063736

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(5), 1621–5.

Hayden, B. Y., & Platt, M. L. (2007). Animal cognition: great apes wait for grapes. *Current biology : CB*, *17*(21), R922–3. doi:10.1016/j.cub.2007.08.061

Hayes, K., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society*.

Haynes, J.-D. (2009). Decoding visual consciousness from human brain signals. *Trends in cognitive sciences*, *13*(5), 194–202. doi:10.1016/j.tics.2009.02.004

Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current biology : CB*, *15*(14), 1301–7. doi:10.1016/j.cub.2005.06.026

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current biology : CB*, *17*(4), 323–8. doi:10.1016/j.cub.2006.11.072

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science (New York, N.Y.)*, *317*(5843), 1360–6. doi:10.1126/science.1146282

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2012). Direct and Indirect Reputation Formation in Nonhuman Great Apes and Human Children. *Journal of comparative psychology (Washington, D.C. : 1983)*.

Herrnstein, R. (1976). Natural concepts in pigeons. *Journal of Experimental* ….

Herrnstein, R., & Loveland, D. (1964). Complex visual concept in the pigeon. *Science*.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Howard, J. D., Plailly, J., Grueschow, M., Haynes, J.-D., & Gottfried, J. A. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nature neuroscience*, *12*(7), 932–8. doi:10.1038/nn.2324

Humphrey, N. (2012a). "This chimp will kick your ass at memory games – but how the hell does he do it?" *Trends in Cognitive Sciences*, *16*(7), 353–355. doi:10.1016/j.tics.2012.05.002

Humphrey, N. (2012b). "This chimp will kick your ass at memory games - but how the hell does he do it?". *Trends in cognitive sciences*, *16*(7), 353–5.

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), 863–6. doi:10.1126/science.1117593

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current biology : CB*, *17*(23), R1004–5. doi:10.1016/j.cub.2007.10.027

Isbell, L. A. (2006). Snakes as agents of evolutionary change in primate brains. *Journal of human evolution*, *51*(1), 1–35. doi:10.1016/j.jhevol.2005.12.012

Janssen, P., Vogels, R., & Orban, G. A. (2000). Three-dimensional shape coding in inferior temporal cortex. *Neuron*, *27*(2), 385–97.

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

Johnston, R. E., & Peng, A. (2008). Memory for individuals: hamsters (Mesocricetus auratus) require contact to develop multicomponent representations (concepts) of others. *Journal of comparative psychology (Washington, D.C. : 1983)*, *122*(2), 121–31. doi:10.1037/0735-7036.122.2.121

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. doi:10.1167/8.13.11

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: fast categorization and object interference. *Vision research*, *47*(26), 3286–97. doi:10.1016/j.visres.2007.09.013

Gallup J, G, (1968). Mirror-image stimulation. *Psychological Bulletin*.

Kalaska, J. F., & Crammond, D. J. (1992). Cerebral cortical mechanisms of reaching movements. *Science (New York, N.Y.)*, *255*(5051), 1517–23.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679–85. doi:10.1038/nn1444

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current biology : CB*, *16*(11), 1096–102. doi:10.1016/j.cub.2006.04.003

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), 4302–11.

Kanwisher, Nancy. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11163–70. doi:10.1073/pnas.1005062107

Katzner, S., Nauhaus, I., Benucci, A., Bonin, V., Ringach, D. L., & Carandini, M. (2009). Local origin of field potentials in visual cortex. *Neuron*, *61*(1), 35–41. doi:10.1016/j.neuron.2008.11.016

Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima Islet. *Primates*.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–5. doi:10.1038/nature06713

Kellogg, W., & Kellogg, L. (1933). The ape and the child: a study of environmental influence upon early behavior.

Kendrick, K., Atkins, K., & Hinton, M. (1995). Facial and vocal discrimination in sheep. *Animal behaviour*.

Kendrick, K. M., da Costa, A. P., Leigh, A. E., Hinton, M. R., & Peirce, J. W. (2001). Sheep don't forget a face. *Nature*, *414*(6860), 165–6. doi:10.1038/35102669

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision research*, *46*(11), 1762–76. doi:10.1016/j.visres.2005.10.002

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of neurophysiology*, *71*(3), 856–67.

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(7), 2488–92. doi:10.1523/JNEUROSCI.3074-10.2011

Kreiman, G, Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature neuroscience*, *3*(9), 946–53. doi:10.1038/78868

Kreiman, Gabriel, 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), 433–45. doi:10.1016/j.neuron.2005.12.019

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., et al. (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

Ku, S.-P., Tolias, A. S., Logothetis, N. K., & Goense, J. (2011). fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron*, *70*(2), 352–62. doi:10.1016/j.neuron.2011.02.048

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences*, *23*(11), 571–9.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nature reviews. Neuroscience*, *3*(11), 884–95. doi:10.1038/nrn964

Lawick-Goodall, J. V. (1970). Tool-using in primates and other vertebrates. *Advances in the Study of Behavior*.

Lazareva, O. F., Freiburger, K. L., & Wasserman, E. A. (2004). Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychonomic bulletin & review*, *11*(6), 1111–7.

LeDoux, J. (1996). Emotional networks and motor control: a fearful view. *Progress in brain research*, *107*, 437–46.

Lestel, D. (2003). Les origines animales de la culture.

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*(14), 9596–601. doi:10.1073/pnas.092277599

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), 281–90. doi:10.1016/j.neuron.2009.02.025

Lobue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: attention to fear-relevant stimuli by adults and young children. *Psychological science*, *19*(3), 284–9. doi:10.1111/j.1467-9280.2008.02081.x

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *353*(1377), 1801–18. doi:10.1098/rstb.1998.0333

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–7. doi:10.1038/35084005

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, *19*, 577–621. doi:10.1146/annurev.ne.19.030196.003045

Lotte, F., Congedo, M., Lécuyer, a, Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, *4*(2), R1–R13. doi:10.1088/1741-2560/4/2/R01

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. *Animal cognition*, *13*(3), 405–18. doi:10.1007/s10071-009-0290-4

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

Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *The European journal of neuroscience*, *21*(7), 2007–18. doi:10.1111/j.1460-9568.2005.04029.x

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397–405. doi:10.1016/j.neuron.2009.07.012

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in cognitive sciences*, *15*(3), 97–103. doi:10.1016/j.tics.2011.01.004

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., et al. (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), 8135–9.

Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., Orban, G. A., et al. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nature methods*, *9*(3), 277–82. doi:10.1038/nmeth.1868

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*.

Marr, D. (1982). Vision: A computational approach.

Martin, A, Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(6566), 649–52. doi:10.1038/379649a0

Martin, Alex. (2007). The representation of object concepts in the brain. *Annual review of psychology*, *58*, 25–45. doi:10.1146/annurev.psych.57.102904.190143

Martin-Malivel, J, & Fagot, J. (2001). Cross-modal integration and conceptual categorization in baboons. *Behavioural brain research*, *122*(2), 209–13.

Martin-Malivel, Julie, Mangini, M. C., Fagot, J., & Biederman, I. (2006). Do humans and baboons use the same information when categorizing human and baboon faces? *Psychological science*, *17*(7), 599–607. doi:10.1111/j.1467-9280.2006.01751.x

Martin-malivel, J., Mangini, M. C., & Biederman, I. (2006). Do Humans and Baboons Use the Same Information When Categorizing Human and Baboon Faces ?, *17*(7), 599–607.

Matsuzawa, T. (n.d.). Use of numbers by a chimpanzee. *Nature*, *315*(6014), 57–9.

Matsuzawa, Tetsuro. (2009). The chimpanzee mind: in search of the evolutionary roots of the human mind. *Animal cognition*, *12 Suppl 1*, S1–9. doi:10.1007/s10071-009-0277-1

Mauck, B., & Dehnhardt, G. (2005). Identity concept formation during visual multiple-choice matching in a harbor seal (Phoca vitulina). *Learning & behavior*, *33*(4), 428–36.

McComb, K., Baker, L., & Moss, C. (2006). African elephants show high levels of interest in the skulls and ivory of their own species. *Biology letters*, *2*(1), 26–8. doi:10.1098/rsbl.2005.0400

Medin, D. L. (1989). Concepts and conceptual structure. *The American psychologist*, *44*(12), 1469–81.

Mehring, C., Rickert, J., Vaadia, E., Cardosa de Oliveira, S., Aertsen, A., & Rotter, S. (2003). Inference of hand movements from local field potentials in monkey motor cortex. *Nature neuroscience*, *6*(12), 1253–4. doi:10.1038/nn1158

Menzel, E. (1973). Chimpanzee spatial memory organization. *Science*.

Mercado, E., Killebrew, D., Pack, A., Mácha, I., & Herman, L. (2000). Generalization of "same-different" classification abilities in bottlenosed dolphins. *Behavioural processes*, *50*(2-3), 79–94.

Mervis, B., Gray, W. D., & Boyes-braem, P. (1976). in Natural Categories, *439*, 382–439.

Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron*, *66*(4), 501–7. doi:10.1016/j.neuron.2010.04.010

Mirabella, G., Bertini, G., Samengo, I., Kilavik, B. E., Frilli, D., Della Libera, C., & Chelazzi, L. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron*, *54*(2), 303–18. doi:10.1016/j.neuron.2007.04.007

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H. C., Sadato, N., et al. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, *60*(5), 915–29. doi:10.1016/j.neuron.2008.11.004

Monosov, I. E., Trageser, J. C., & Thompson, K. G. (2008). Measurements of simultaneously recorded spiking activity and local field potentials suggest that spatial selection emerges in the frontal eye field. *Neuron*, *57*(4), 614–25. doi:10.1016/j.neuron.2007.12.030

Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., et al. (2011). A category-specific response to animals in the right human amygdala. *Nature neuroscience*, *14*(10), 1247–9. doi:10.1038/nn.2899

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Context familiarity enhances target processing by inferior temporal cortex neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(32), 8533–45. doi:10.1523/JNEUROSCI.2106-07.2007

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological cybernetics*, *65*(2), 135–45.

Munksgaard, L., De Passillé, A. M., Rushen, J., Thodberg, K., & Jensen, M. B. (1997). Discrimination of people by dairy cows based on handling. *Journal of dairy science*, *80*(6), 1106–12. doi:10.3168/jds.S0022-0302(97)76036-3

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–15. doi:10.1016/j.neuron.2009.09.006

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(39), 13771–85. doi:10.1523/JNEUROSCI.2792-11.2011

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology : CB*, *21*(19), 1641–6. doi:10.1016/j.cub.2011.08.031

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of experimental psychology. General*, *130*(3), 466–78.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review*, *108*(3), 483–522.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature reviews. Neuroscience*, *9*(2), 123–35. doi:10.1038/nrn2314

Orban, G. A. (2011). The extraction of 3D shape in the visual system of human and nonhuman primates. *Annual review of neuroscience*, *34*(March), 361–88. doi:10.1146/annurev-neuro-061010-113819

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), 315–24. doi:10.1016/j.tics.2004.05.009

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic representation of the human body in the occipitotemporal cortex. *Neuron*, *68*(3), 586–600. doi:10.1016/j.neuron.2010.09.032

Paradiso, M. A., Blau, S., Huang, X., MacEvoy, S. P., Rossi, A. F., & Shalev, G. (2006). Lightness, filling-in, and the fundamental role of context in visual perception. *Progress in brain research*, *155*, 109–23. doi:10.1016/S0079-6123(06)55007-1

Parron, C., Call, J., & Fagot, J. (2008). Behavioural responses to photographs by pictorially naïve baboons (Papio anubis), gorillas (Gorilla gorilla) and chimpanzees (Pan troglodytes). *Behavioural processes*, *78*(3), 351–7. doi:10.1016/j.beproc.2008.01.019

Parron, C., & Fagot, J. (2008). Baboons (Papio papio) spontaneously process the first-order but not second-order configural properties of faces. *American journal of primatology*, *70*(5), 415–22. doi:10.1002/ajp.20503

Pascalis, O, & Bachevalier, J. (1998). Face recognition in primates: a cross-species study. *Behavioural Processes*.

Pascalis, Olivier, de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science (New York, N.Y.)*, *296*(5571), 1321–3. doi:10.1126/science.1070223

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS biology*, *10*(1), e1001251. doi:10.1371/journal.pbio.1001251

Pasupathy, A, & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of neurophysiology*, *82*(5), 2490–502.

Pasupathy, Anitha. (2006). Neural basis of shape representation in the primate brain. *Progress in brain research*, *154*, 293–313. doi:10.1016/S0079-6123(06)54016-6

Pasupathy, Anitha, & Connor, C. E. (2002). Population coding of shape in area V4. *Nature neuroscience*, *5*(12), 1332–8. doi:10.1038/nn972

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(30), 10127–34. doi:10.1523/JNEUROSCI.2161-10.2010

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–7. doi:10.1038/nature08103

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125–30. doi:10.1073/pnas.1101042108

Peirce, J. W., Leigh, A. E., & Kendrick, K. M. (2000). Configurational coding, familiarity and the right hemisphere advantage for face recognition in sheep. *Neuropsychologia*, *38*(4), 475–83.

Peirce, J. W., Leigh, A. E., daCosta, A. P. C., & Kendrick, K. M. (2001). Human face recognition in sheep: lack of configurational coding and right hemisphere advantage. *Behavioural processes*, *55*(1), 13–26.

Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(6), 1323–34. doi:10.1093/cercor/bhl043

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *The Behavioral and brain sciences*, *31*(2), 109–30; discussion 130–178. doi:10.1017/S0140525X08003543

Perry, E., & Renouf, D. (1988). Further studies of the role of harbour seal (Phoca vitulina) pup vocalizations in preventing separation of mother-pup pairs. *Canadian journal of zoology*.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. *Nature reviews. Neuroscience*, *11*(11), 773–83. doi:10.1038/nrn2920

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, 2581–2600. doi:10.1152/jn.91198.2008.

Plailly, J., Demarquay, G., & Royet, J. P. (2009). Cerebral imaging and olfactory disorders: a review. *B-ENT*, *5 Suppl 13*, 61–9.

Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(45), 17053–7. doi:10.1073/pnas.0608062103

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science (New York, N.Y.)*, *310*(5756), 1963–6. doi:10.1126/science.1117645

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

Premack, D. (1976). Intelligence in ape and man.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*.

Premack, D. (2010a). Why Humans Are Unique Three Theories. *Perspectives on Psychological Science*.

Premack, D. (2010b). Why Humans Are Unique: Three Theories. *Perspectives on Psychological Science*, *5*(1), 22–32. doi:10.1177/1745691609356782

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*.

Press, C., & Sciences, B. (2012). Do Animals Have Personality? *BioScience*, *62*(6), 533–537. doi:10.1525/bio.2012.62.6.4

Preuss, T. M. (2012). Human brain evolution: From gene discovery to phenotype discovery. *Proceedings of the National Academy of Sciences of the United States of America*, *109 Suppl* , 10709–16. doi:10.1073/pnas.1201894109

Principle, T. (1977). Forest Before Trees : The Precedence of Global in Visual Perception, *383*, 353–383.

Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in domestic horses (Equus caballus). *Proceedings of the National Academy of Sciences of the United States of America*, *106*(3), 947–51. doi:10.1073/pnas.0809127105

Proust, J. (1997). Comment l'esprit vient aux bêtes: essai sur la représentation.

Quiroga, R. Q., Reddy, L., Koch, C., & Fried, I. (2007). Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of neurophysiology*, *98*(4), 1997–2007. doi:10.1152/jn.00125.2007

Rajimehr, R., Young, J. J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 1995–2000. doi:10.1073/pnas.0807304106

Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal cognition*, *11*(2), 339–47. doi:10.1007/s10071-007-0123-2

Ray, S., & Maunsell, J. H. R. (2010). Differences in gamma frequencies across visual cortex restrict their possible use in computation. *Neuron*, *67*(5), 885–96. doi:10.1016/j.neuron.2010.08.004

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–25. doi:10.1016/j.neuroimage.2009.11.084

Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(10), 5937–42. doi:10.1073/pnas.101086398

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, *2*(11), 1019–25. doi:10.1038/14819

Roberts, W., & Mazmanian, D. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes*.

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, perception & psychophysics*, *73*(8), 2542–72. doi:10.3758/s13414-011-0200-0

Rosati, A. G., Stevens, J. R., Hare, B., & Hauser, M. D. (2007). The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Current biology : CB*, *17*(19), 1663–8. doi:10.1016/j.cub.2007.08.033

Rosch, E., Mervis, C., & Gray, W. (1976). Basic objects in natural categories. *Cognitive* ….

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*(4), 1959–79. doi:10.1016/j.neuroimage.2007.10.011

Rougeul-Buser, A., Bouyer, J. J., & Buser, P. (1975). From attentiveness to sleep. A topographical analysis of localized "synchronized" activities on the cortex of normal cat and monkey. *Acta neurobiologiae experimentalis*, *35*(5-6), 805–19.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature neuroscience*, *5*(7), 629–30. doi:10.1038/nn866

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Spatiotemporal analyses of the N170 for human faces, animal faces and objects in natural scenes. *Neuroreport*, *15*(17), 2607–11.

Rousselet, G. A., Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2007). Limits of event-related potential differences in tracking object processing speed. *Journal of cognitive neuroscience*, *19*(8), 1241–58. doi:10.1162/jocn.2007.19.8.1241

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004a). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision research*, *44*(9), 877–94. doi:10.1016/j.visres.2003.11.014

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004b). How parallel is visual processing in the ventral pathway? *Trends in cognitive sciences*, *8*(8), 363–70. doi:10.1016/j.tics.2004.06.003

Rumbaugh, D., & Pate, J. (1984). The evolution of cognition in primates: A comparative perspective. *Animal cognition*.

Ryabko, B., & Reznikova, Z. (2011). Numerical competence in animals, with an insight from ants. *Behaviour*, *148*(4), 405–434. doi:10.1163/000579511X568562

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of experimental psychology. General*, *111*(4), 369–89.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1978). Symbolic communication between two chimpanzees (Pan troglodytes). *Science (New York, N.Y.)*, *201*(4356), 641–4.

Schell, A., Rieck, K., Schell, K., Hammerschmidt, K., & Fischer, J. (2011). Adult but not juvenile Barbary macaques spontaneously recognize group members from pictures. *Animal cognition*, *14*(4), 503–9. doi:10.1007/s10071-011-0383-8

Schiller, P. H. (1995). Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature*, *376*(6538), 342–4. doi:10.1038/376342a0

Schlag, J., Cai, R. H., Dorfman, A., Mohempour, A., Schlag-rey, M., Bernardo, D., & Murray, A. (2000). Numerical memory span Explaining the T-wave shape in the ECG, 39–41.

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*(15), 6424–9. doi:10.1073/pnas.0700622104

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science (New York, N.Y.)*, *210*(4471), 801–3.

Sheehan, M. J., & Tibbetts, E. A. (2011). Specialized face learning is associated with individual recognition in paper wasps. *Science (New York, N.Y.)*, *334*(6060), 1272–5. doi:10.1126/science.1211334

Shepard, R. (1961). Learning and memorization of classifications. … *Monographs: General and ….*

Shibasaki, M., & Kawai, N. (2009). Rapid detection of snakes by Japanese monkeys (Macaca fuscata): an evolutionarily predisposed visual system. *Journal of comparative psychology (Washington, D.C. : 1983)*, *123*(2), 131–5. doi:10.1037/a0015095

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

Sliwa, J., Duhamel, J.-R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1735–40. doi:10.1073/pnas.1008169108

Smith, J. D. (2009). The study of animal metacognition. *Trends in cognitive sciences*, *13*(9), 389–96. doi:10.1016/j.tics.2009.06.009

Smith, J., Shields, W., Schull, J., & Washburn, D. (1997). The uncertain response in humans and animals. *Cognition*.

Soares, S. C. (2012). The lurking snake in the grass: Interference of snake stimuli in visually taxing conditions. *Evolutionary psychology : an international journal of evolutionary approaches to psychology and behavior*, *10*(2), 187–97.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, *11*(5), 543–5. doi:10.1038/nn.2112

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–8.

Swaminathan, S. K., & Freedman, D. J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature neuroscience*, *15*(2), 315–20. doi:10.1038/nn.3016

Tate, A. J., Fischer, H., Leigh, A. E., & Kendrick, K. M. (2006). Behavioural and neurophysiological evidence for face identity and face emotion processing in animals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *361*(1476), 2155–72. doi:10.1098/rstb.2006.1937

Teller, D. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development*.

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature neuroscience*, *10*(4), 505–11.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, *9*(4), 483–4.

Thompson, R. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, *24*(3), 363–396.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–2. doi:10.1038/381520a0

Thorpe, S. J. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*(5502), 260.

Tibbetts, E. A. (2002). Visual signals of individual identity in the wasp Polistes fuscatus. *Proceedings. Biological sciences / The Royal Society*, *269*(1499), 1423–8. doi:10.1098/rspb.2002.2031

Tolman, E. (1927). A behaviorist's definition of consciousness. *Psychological Review*.

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

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: an update. *Annals of the New York Academy of Sciences*, *1124*, 239–61. doi:10.1196/annals.1440.004

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network (Bristol, England)*, *14*(3), 391–412.

Treisman, A. (1996). The binding problem. *Current opinion in neurobiology*, *6*(2), 171–8.

Tricarico, E., Borrelli, L., Gherardi, F., & Fiorito, G. (2011). I know my neighbour: individual recognition in Octopus vulgaris. *PloS one*, *6*(4), e18710. doi:10.1371/journal.pone.0018710

Truppa, V., Spinozzi, G., Stegagno, T., & Fagot, J. (2009). Picture processing in tufted capuchin monkeys (Cebus apella). *Behavioural processes*, *82*(2), 140–52. doi:10.1016/j.beproc.2009.05.004

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), 989–95. doi:10.1038/nn1111

Tsao, D. Y., Freiwald, W. a, Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science (New York, N.Y.)*, *311*(5761), 670–4. doi:10.1126/science.1119983

Tsao, D. Y., Moeller, S., & Freiwald, W. a. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19514–9. doi:10.1073/pnas.0809662105

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature neuroscience*, *8*(8), 1096–101. doi:10.1038/nn1500

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature neuroscience*, *5*(7), 682–7. doi:10.1038/nn870

Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 883–90.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science (New York, N.Y.)*, *255*(5043), 419–23.

VanRullen, R. (2006). The continuous Wagon Wheel Illusion is object-based. *Vision research*, *46*(24), 4091–5. doi:10.1016/j.visres.2006.07.030

Vanrullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*.

Vauclair, J., & Fagot, J. (1996). Categorization of alphanumeric characters by Guinea baboons: Within-and between-class stimulus discrimination. *Cahiers de Psychologie Cognitive*.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *The European journal of neuroscience*, *11*(4), 1223–38.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in cognitive sciences*, *9*(12), 585–94. doi:10.1016/j.tics.2005.10.011

Waal, F. D. (2007). Chimpanzee politics: Power and sex among apes.

Wasserman, E., Kiedinger, R., & Bhatt, S. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, *Vol 14(3)*, 235–246.

Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. *Neuropsychologia*, *47*(14), 3111–7. doi:10.1016/j.neuropsychologia.2009.07.008

Woodman, G. F., Kang, M.-S., Rossi, A. F., & Schall, J. D. (2007). Nonhuman primate event-related potentials indexing covert shifts of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(38), 15111–6. doi:10.1073/pnas.0703477104

Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: response to mixed visual displays after same-different discrimination training. *Journal of experimental psychology. Animal behavior processes*, *23*(2), 157–70.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*(3), 429–41. doi:10.1016/j.neuron.2008.03.027

Zayan, R., & Vauclair, J. (1998). Categories as paradigms for comparative cognition. *Behavioural Processes*.

LISTE DES PUBLICATIONS

Articles publiés dans des journaux internationaux à comité de lecture

Fize, D., **Cauchoix, M**. & Fabre-Thorpe M. (2011). Humans and monkeys share visual representations. ***PNAS***,108 (18) 7635-7640

Crouzet C., **Cauchoix M**. (2011). When Does the Visual System Need to Look Back? ***J. Neurosci.,*** 31(24): 8706-8707;

Cauchoix, M., Arslan, A. B., Fize, D., & Serre, T. (2011). The neural dynamics of visual processing in monkey extrastriate cortex : A comparison between univariate and multivariate techniques. (Sous presse). ***Lecture Notes in Computer Science***

Articles soumis (ou prêts pour soumission)

Cauchoix M, Crouzet S, Serre T, Fize D, Early neural activity from the visual ventral stream enable rapid categorization

Cauchoix M. Crouzet SM. How plausible is a subcortical account of rapid visual recognition?

Cauchoix M., Barragan-Jason G., Serre T., Barbeau E., Fast EEG decoding of human faces in the wild

Barragan-Jason G.**, Cauchoix M**., Barbeau E The time course of famous face recognition

Articles en preparation

Cauchoix M, Fabre-Thorpe M Utra-fast cortical selectivity for snakes in macaque visual ventral stream

Barragan-Jason G., **Cauchoix M.**, Crouzet S.M., Barbeau E.J The two facets of face processing: insights from behavioral and electrophysiological studies

Barragan-Jason G., **Cauchoix M.,** Valton L., Sol J.C., Barbeau E. The neuronal dynamics of face processing acess form intracranial recording in human

Cauchoix M., Barragan-Jason G.,Valton L., Sol J.C., Barbeau E. Timing of natural categories and task status encoding in human intracranial recordings

Présentations orales

Cauchoix, M., Fabre-Thorpe, M., & Fize, D. (2011) Shared visual representations in humans and monkeys. Perception (Vol. 40, p. 32), ECVP, Toulouse, France.

Cauchoix, M., Serre, T., Kreiman,G., Fize, D. (2010) Fast decoding of natural object categories from intracranial field potentials in monkey's visual cortex. Alpine Brain Imaging Meeting, Champery, Switzerland

Cauchoix, M., Serre, T., Kreiman,G., Fize, D. (2010) Fast decoding of natural object categories from intracranial field potentials in monkey's visual cortex. Journal of Vision, 10(7): 947, VSS, Naples, Florida, USA.

Cauchoix, M., Fabre-Thorpe, M., Fize, D. (2009) Forest before Trees in Natural Scenes: Behavioural and Neural evidence in Monkeys. Thirteen International Conference on Cognitive and Neural Systems, Boston, USA.

Présentations affichées

Cauchoix, M., Fabre-Thorpe, M., & Fize, D. (2009) Dealing with natural scenes: Electrocorticographic correlates of object and context processing in Rhesus Monkey. Journal of Vision, 9(8):960. VSS, Naples, Florida, USA.

Cauchoix M., Crouzet S., Fize D., Serre T. (2011) Visual features and dynamics of rapid recognition in the visual cortex. Soc. Neurosc. Abstr., Washington, USA.

Crouzet, S., **Cauchoix**, M., Fize, D., & Serre, T. (2011). The neural basis of rapid categorization : Linking computational models and electrophysiology. Advances in neural information processing, NIPS 2012, Granada, Spain

Arslan, A., **Cauchoix**, M., Singer, J., Fize, D., Kreiman, G., & Serre, T. (2011) The neural basis of rapid visual recognition: Neural decoding and Granger causality analysis of connectivity. Journal of Vision, 11(11):824. VSS, Naples, Florida, USA.

Résumé

Est-ce que les animaux pensent ? Quelle est la nature des mécanismes qui dirigent leurs comportements? Peut-on comparer ces mécanismes au fonctionnement de l'esprit humain ? Ces questions animent les débats scientifiques, philosophiques et théologiques depuis des millénaires. L'idée d'une discontinuité fondamentale entre l'esprit humain et les rouages de l'animal-machine, prônée par Descartes, s'oppose à l'hypothèse de continuité de Darwin selon laquelle, la différence entre les capacités mentales de l'homme et de l'animal serait de degré et non de nature. Dans cette thèse, fort des récentes avancées méthodologiques et théoriques dans l'étude du comportement animal et des neurosciences, nous partons à la recherche de la pierre de rosette pouvant éclairer ce débat : l'homologie neuro-cognitive fonctionnelle. Existe-t-il un fonctionnement nerveux commun à plusieurs espèces (et notamment à l'homme), supportant un même fonctionnement cognitif ?

Nous focalisons notre recherche sur l'une des fonctions clefs de l'esprit humain, qui semble aussi très répandue dans le monde animal : la catégorisation. Former des concepts -regrouper mentalement les objets ou événements auxquelles nous sommes confrontés en catégories- permet de simplifier, de donner un sens et de réagir de manière appropriée aux flux perceptuels incessants issus de nos sens.

En testant des sujets humains et macaques dans un protocole rigoureusement identique de catégorisation visuelle rapide, en particulier, en contrôlant l'influence de l'information contextuelle de la scène visuelle sur les performances de catégorisation de l'objet, nous montrons ici que les deux espèces de primates semblent bien avoir recours à des représentations ou concepts visuels similaires pour résoudre cette tâche. Mieux, le développement de nouvelles méthodes d'analyses multivariées des potentiels de champs corticaux, nous permet de lire, essai par essai, l'émergence de représentations neuronales associées à cette tâche de catégorisation. En appliquant ces méthodes chez le singe, en EEG de surface chez l'homme et chez des patients épileptiques implantés pour raison médicale, on met en évidence le rôle analogue, chez ces deux espèces de primates, de la voie visuelle ventrale dans la formation ultra-rapide de ces représentations catégorielles.

Pour résumer, la capacité de catégorisation visuelle rapide, semble être un trait neurocognitif partagé par l'homme et le macaque, probablement héritée d'un ancêtre commun. L'existence d'une telle homologie est un argument pour l'hypothèse de continuité et permet donc d'envisager une évolution de la cognition. Enfin, si on considère la catégorisation comme un élément clef de la pensée humaine, alors ces résultats suggèrent l'existence d'une pensée chez le macaque, dont au moins une partie fonctionne comme la nôtre.

Abstract

Do animals think? What are the mechanisms supporting their behavior? Can we compare these mechanisms with human mind? These questions animate scientific, philosophical and theological debates for millennia. The idea of a fundamental discontinuity between human mind and animal-machine workings, advocated by Descartes, is opposed to Darwin's assumption of continuity that suppose a difference of degree and not of kind between human's and animals' minds. In this thesis, based on recent theoretical and methodological advances in animal behavior and neuroscience, we go in search of the Rosetta Stone that can inform this debate: a functional neuro-cognitive homology. Is there a neural mechanism, shared by several species (including humans), supporting the same cognitive function?

We focus our research on one of the key functions of human mind, which also seems widespread in the animal world: categorization ability. The making of concepts –to mentally regroup objects, or events, we face in categories- simplifies, makes sense and allows responding appropriately to the continuous perceptual streams. Testing humans and macaques in exactly the same protocol of rapid visual categorization, in particular, by controlling the influence of contextual information from the visual scene on object categorization performances, we show here that the two primates seem to rely on similar concepts or visual representations to solve this task. More, with a new multivariate pattern analysis for cortical field potentials, we can read, trial by trial, the emergence of neural representations associated with such categorization task. By applying these methods in monkeys, surface EEG in humans and in patients with epilepsy, implanted for medical reasons, we highlight the similar role of the ventral visual pathway to build quick categorical representations, in both species of primates.

To summarize, the ability of rapid visual categorization seems to be a neuro-cognitive trait shared by humans and macaques, probably inherited from a common ancestor. The existence of such homology is an argument for the continuity hypothesis and can therefore justify an evolution of cognition. Finally, if we consider the categorization as a key element of human thought, then these results suggest the existence of a thought in macaques, which at least partially works like ours.