Université de Montréal

Décoder l'habileté perceptive dans le cerveau humain :

contenu représentationnel et computations cérébrales

Par

Simon Faghel-Soubeyrand

Département de Psychologie, Faculté des Arts et des Sciences

Thèse présentée en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.) en Psychologie,
Option Neuroscience Cognitive

25 novembre 2022

Université de Montréal

Unité académique : département de Psychologie, Faculté des Arts et des Sciences

---

Cette thèse intitulée

Décoder l'habileté perceptive dans le cerveau humain :

contenu représentationnel et computations cérébrales

Présenté par

Simon Faghel-Soubeyrand

A été évaluée par un jury composé des personnes suivantes

Hugo Théorêt
Président-rapporteur

Frédéric Gosselin
Directeur de recherche

Ian Charest
Codirecteur

Shahab Bakhtiari
Membre du jury

Leyla Isik
Examinateur externe

Résumé

La capacité à reconnaître les visages de nos collègues, de nos amis et de nos proches est essentielle à notre réussite en tant qu'êtres sociaux. Notre cerveau accomplit cet exploit facilement et rapidement, dans une série d'opérations se déroulant en quelques dizaines de millisecondes à travers un vaste réseau cérébral du système visuel ventral. L'habileté à reconnaître les visages, par contre, varie considérablement d'une personne à l'autre. Certains individus, appelés «*super-recognisers*», sont capables de reconnaître des visages vus une seule fois dans la rue des années plus tôt. D'autres, appelés «prosopagnosiques», sont incapables de reconnaître le visage de leurs collègues ou leurs proches, même avec une vision parfaite. Une question simple reste encore largement sans réponse : *quels mécanismes expliquent que certains individus sont meilleurs à reconnaître des visages?* Cette thèse rapporte cinq articles étudiant les mécanismes perceptifs (articles 1, 2, 3) et cérébraux (articles 4, 5) derrière ces variations à travers différentes populations d'individus.

L'**article 1** décrit le contenu des représentations visuelles faciales chez une population avec un diagnostic de schizophrénie et d'anxiété sociale à l'aide d'une technique psychophysique *Bubbles*. Nous révélons pour la première fois les mécanismes en reconnaissance des expressions de cette population: un déficit de reconnaissance est accompagné par i) une sous-utilisation de la région des yeux des visages expressifs et ii) une sous-utilisation des détails fins. L'**article 2** valide ensuite une nouvelle technique permettant de révéler simultanément le contenu visuel dans trois dimensions psychophysiques centrales pour le système visuel — la position, les fréquences spatiales, et l'orientation. L'**article 3** a mesuré, à l'aide de cette nouvelle technique, le contenu représentationnel de 120 individus pendant la discrimination faciale du sexe et des expressions (>500,000 observations). Nous avons observé de fortes corrélations entre l'habileté à discriminer le sexe et les expressions des visages, ainsi qu'entre l'habileté à discriminer le sexe et l'identité. Crucialement, plus un individu est habile en reconnaissance faciale, plus il utilise un contenu représentationnel similaire entre les tâches.

L'**article 4** a examiné les computations cérébrales de *super-recognisers* en utilisant l'électroencéphalographie haute-densité (EEG) et l'apprentissage automatique. Ces outils ont permis de décoder, pour la première fois, l'habileté en reconnaissance faciale à partir du cerveau avec jusqu'à 80% d'exactitude — et ce à partir d'une seule seconde d'activité cérébrale. Nous avons ensuite utilisé la Representational Similarity Analysis (RSA) pour comparer les représentations cérébrales de nos participants à celles de modèles d'apprentissage profond visuels et langagiers. Les *super-recognisers*, comparé aux individus avec une habileté typique, ont des représentations cérébrales plus similaires aux computations visuelles et sémantiques de ces modèles optimaux. L**'article 5** rapporte une investigation des computations cérébrales chez le cas le plus spécifique et documenté de *prosopagnosie acquise*, la patiente PS. Les mêmes outils computationnels et d'imagerie que ceux de l'article 4 ont permis i) de décoder les déficits d'identification faciale de PS à partir de son activité cérébrale EEG, et ii) de montrer pour la première fois que la prosopagnosie est associée à un déficit des computations visuelles de haut niveau et des computations cérébrales sémantiques.

Mots-clés : Vision, reconnaissance faciale, différences individuelles, psychophysique, électroencéphalographie, RSA, apprentissage automatique, apprentissage profond, prosopagnosie, super-recognisers.

Abstract

The ability to recognise the faces of our colleagues, friends, and family members is critical to our success as social beings. Our brains accomplish this feat with astonishing ease and speed, in a series of operations taking place in tens of milliseconds across a vast brain network of the visual system. The ability to recognise faces, however, varies considerably from one person to another. Some individuals, called "super-recognisers", are able to recognise faces seen only once years earlier. Others, called "prosopagnosics", are unable to recognise the faces of their colleagues or relatives, even with perfect vision and typical intelligence. A simple question remains largely unanswered: *what mechanisms explain why some individuals are better at recognizing faces*? This thesis reports five articles studying the perceptual (article 1, 2, 3) and neural (article 4, 5) mechanisms behind these variations across different populations of individuals.

Article 1 describes the content of visual representations of faces in a population with a comorbid diagnosis of schizophrenia and social anxiety disorder using an established psychophysical technique, *Bubbles*. We reveal for the first time the perceptual mechanisms of expression recognition in this population: a recognition deficit is accompanied by i) an underutilization of the eye region of expressive faces and ii) an underutilization of fine details. **Article 2** then validates a new psychophysical technique that simultaneously reveals the visual content in three dimensions central to the visual system — position, spatial frequencies, and orientation. We do not know, however, whether skilled individuals perform well across a variety of facial recognition tasks and, if so, how they accomplish this feat. **Article 3** measured, using the technique validated in article 2, the perceptual representations of 120 individuals during facial discrimination of gender and expressions (total of >500,000 trials). We observed strong correlations between the ability to discriminate gender and facial expressions, as well as between the ability to discriminate gender and identify faces. More importantly, we found a positive correlation between individual ability and the similarity of perceptual representations used across these tasks.

Article 4 examined differences in brain dynamics between super-recognizers and typical individuals using high-density electroencephalography (EEG) and machine learning. These tools allowed us to decode, for the first time, facial recognition ability from the brain with up to 80% accuracy — using a mere second of brain activity. We then used Representational Similarity Analysis (RSA) to compare our participants' brain representations to those of deep learning models of object and language classification. This showed that super-recognisers, compared to individuals with typical perceptual abilites, had brain representations more similar to the visual and semantic computations of these optimal models. **Article 5** reports an investigation of brain computations in the most specific and documented case of acquired prosopagnosia, patient PS. The same computational tools used in article 4 enabled us to decode PS's facial identification deficits from her brain dynamics. Crucially, associations between brain deep learning models showed for the first time that prosopagnosia is associated with deficits in high-level visual and semantic brain computations.

Keywords: vision, facial recognition, individual differences, psychophysics, electroencephalography, RSA, machine learning, deep learning, prosopagnosia, super-recognizers

Table des matières

Liste des tableaux

Chapitre 2 : Article 1

Liste des Figures

Chapitre 1 : Introduction générale

Chapitre 2 : Article 1

Chapitre 3 : Article 2

Chapitre 4 : Article 3

Chapitre 5 : Article 4

Chapitre 6 : Article 5

Chapitre 7 : Discussion générale

Liste des sigles et abréviations

ATL : lobe temporal antérieure

CNN : convolutional neural networks

CGL : Corps Genouillé Latéral

CI : Classification Image

dl-PFC : cortex préfrontal dorso-latéral

ERP : *Event Related Potentials*

EEG : électroencéphalographie

FS : fréquences spatiales

FFA : Fusiform Face Area

GUSE : *Google Universal Sentence Encoder*

hIT: human infero temporal cortex

IM : Intermodular frequencies

IT : infero temporal cortex

IRMf: Imagerie par résonance magnétique fonctionelle

LDA : Linear Discriminant

LOC: Lateral Occipital Cortex

MI : Mutual Information

MTL : lobe temporal médian

NCC : neurosciences cognitives computationelles

OFA: Occipital Face Area

pCC: posterior cingulate cortex

PPA: Parahipocampal Place Area

PET : *Position Emission Tomography*

RSA : *Representational Similarity Analysis*

RDM : *Representational Dissimilarity Matrix*

SAD : *Social Anxiety Disorder*

SSVEP : *Steady-State Visual Evoked Potentials*

SZ&SAD : s*chizophrenia & social anxiety disorder*

SZ : *schizophrenia*

Remerciements

Je me rends compte, avec le recul que m'apporte cet ouvrage, du nombre d'individus m'ont aidé, influencé, et supporté durant les 5 ans qui se sont écoulées depuis le début ma thèse. Certains me viennent à l'esprit tout naturellement.

Merci, d'abord, à mes directeurs de thèse, Frédéric et Ian. Frédéric, on se connaît depuis maintenant près de 8 ans. 8 ans, entre autres, à s'opiniâtrer sur des sujets de société. Sans blague, ces 8 années ont été remplies de discussions riches, foisonnantes, et créatives (et, oui, quelques fois intenses) que j'ai toujours énormément appréciées. Merci pour tous ces moments. Je me dois aussi de te remercier pour ton accessibilité, ta compréhension, et tes conseils tout au long de ces cinqs années de PhD. Ton support et l'influence que tu as eu sur ma vie au laboratoire pendant ces années m'ont beaucoup apporté.

Un énorme merci à Ian. Merci pour ta présence, qui depuis 5 ans m'a permit de m'ouvrir et d'explorer un peu plus le monde, géographiquement et scientifiquement. Merci de m'avoir fait découvrir la magnifique et ensoleillée Brum et de m'avoir aidé à m'y faire une place. Merci, plus généralement, pour ton accessibilité, ton accueil, et ta générosité de toute heure. Et, bien sûr : merci pour le Blazer. Fier d'avoir mes initiales à côté des tiennes.

Merci à vous deux d'avoir été une influence positive sur ce qui fait de moi le chercheur que je suis aujourd'hui. Si ces dernières années ont été riches en changement et en développement, c'est entre autres grâce à vous.

Merci à tous mes collègues et ami(e)s du laboratoire. Merci notamment à Laurent et Jessica, qui ont été là depuis le début de mon M.Sc. et avec qui j'ai eu la chance de partager beaucoup de bons moments, allant des conférences, aux sorties de labs, aux nombreuses pauses cafés au bureau et autres discussions de cadre de porte. Ces moments et ces années passées au laboratoire avec vous m'ont beaucoup apporté, et ont certainement participé à ma motivation à continuer en recherche. J'inclus dans ces «vieux» camarades Nicolas Dupuis, qui m'a supervisé à mes tout débuts dans le laboratoire et qui m'a énormément appris. Merci pour ton écoute, ta patience et ta didactique. Merci également à ceux du labo que j'ai rencontrés plus récemment, mais avec qui j'ai tout autant apprécié passer ces moments de travail et de conférences, entre autre: Peter, Catherine, Jean-Maxime, Clémentine, et Michèle. Merci à vous tous.tes d'avoir aidé à passer au dessus de l'isolement post-pandémie, et d'avoir enduré mes discours sur les situations inéquitables des étudiants. Je suis chanceux d'avoir pu finir ce doctorat aussi bien entouré. Une mention également aux bonnes âmes affutées et accueillantes rencontrées hors Montréal; Adrien Doerig, Walter Setti, Luca, Mircea et Eden, entre autre, Ludwig Barbaro, Jasper Van Den Bosch

du côté de Birmingham, et du côté de Fribourg, Camille Saumure, Anne-Raphaelle Richoz, et Roberto Caldara. Merci, d'ailleurs, Roberto, pour ton support répété.

Merci, évidemment, à tous mes proches et ami(e)s de l'autre vie à Montréal ; je n'y serais pas arrivé sans les soirées qui ne se terminent pas, les discussions profondes, et d'autres discussions/moments plus immatures, mais tout aussi agréables. Merci pour votre ouverture et votre curiosité. Vous aurez su faire décrocher, pour le meilleur de ma santé mentale, un esprit qui a de la difficulté à le faire. Un merci particulier à Catherine, qui a vécu (et parfois subit) le début de ce Doctorat avec moi, entre Montréal et Birmingham.

Merci, finalement, à ma Frédérique. Je te l'ai déjà dit, mais je ne pourrai jamais assez te remercier pour tout ce que tu m'as apporté ces deux dernières années. Je ne sais sincèrement pas ce que j'aurais fais sans ta joie de vivre contagieuse, ton écoute, et ton optimisme (même s'il peut être exagéré). Merci pour ton ouverture et ton écoute sur mon monde parfois opaque, pour ton aide constante et quotidienne, et pour ton support inconditionnel dans les moments plus difficiles. J'espère que tu sais que ça n'aurait pas été pareil sans toi.

Donc merci tout le monde. La qualité de vos interactions et l'influence que vous avez tous.tes eus sur cet ouvrage, mais aussi sur moi en tant que personnes durant ces 5 années, est inestimable.

Chapitre 1 – Introduction Générale

1.1 Présentation de la thèse

Avez-vous un talent particulier pour reconnaître les visages des gens de votre entourage? Ou au contraire, avez-vous de la difficulté à suivre l'histoire d'un film parce que vous confondez les acteur.rice.s? La capacité à reconnaître les visages de nos collègues, de nos amis et de nos proches est essentielle à notre quotidien. Notre cerveau accomplit cet exploit avec une facilité et une rapidité étonnante, dans une série d'opérations se déroulant en quelques dizaines de millisecondes à travers un vaste réseau cérébral du système visuel ventral. L'habileté à reconnaître les visages, par contre, varie considérablement d'une personne à l'autre. Certains individus, appelés «*super-recognisers*» ou «super-physionomistes», sont capables de reconnaître des visages vus une seule fois dans la rue des années plus tôt. D'autres, appelés « prosopagnosiques», sont incapables de reconnaître le visage de leurs collègues, leurs ami.e.s ou leurs proches, même avec une vision parfaite et une intelligence typique. Une question simple reste encore largement sans réponse : *quels mécanismes expliquent que certains individus sont meilleurs à reconnaître des visages?*

Différents facteurs peuvent intuitivement influencer cette habileté. Pensez, par exemple, au dernier visage que vous avez croisé. Plusieurs types d'information émergent naturellement: la forme et la couleur de ses yeux, la position et la taille de son nez par rapport à ceux-ci, mais possiblement aussi de l'information sur son identité, son genre, son humeur, et peut-être même l'emploi qu'il/elle occupe. Notre cerveau (re)construit ces différentes représentations à travers de multiples étapes de traitement dans la hiérarchie du système cérébrale visuel, ainsi qu'à travers des régions cérébrales plus antérieures permettant l'émergeance d'informations plus abstraites. Quelle information visuelle spécifique (e.g. couleur, détails fins/grossiers, orientation) distingue la représentation d'un individu à un autre, et quelle partie de ce *contenu représentationnel* informe nos prises de décisions quotidiennes? Nous savons maintenant que des différences importantes dans ce contenu représentationnel existent entre individus, et que ces différences dans le contenu visuel d'une représentation faciale ne sont pas aléatoires. Elles sont même directement liées à l'habileté perceptive : plus un individu utilise les yeux dans sa représentation d'un visage, par

exemple, meilleur il sera pour reconnaître son identité. Le contenu visuel facial dépend, par contre, de la tâche à effectué. En ce moment, nous n'avons aucune idée de l'influence de l'habileté sur les variations du contenu visuel *entre les tâches*. Nous ne savons pas, par exemple, si les individus habiles ont tendance à utiliser la même stratégie quand ils reconnaissent le genre que quand ils reconnaissent l'identité d'un visage.

Les différences d'habileté perceptive émergent aussi nécessairement de variations dans les processus cérébraux à différents niveaux de cette hiérarchie de traitement. En ce moment, nous commençons à peine à savoir *où*, dans le cerveau, les représentations de visages de se forment. Le peu d'études en imagerie cérébrale ayant étudié l'habileté perceptive, par exemple, a révélé que des différences interindividuelles dans l'activation de certaines régions du système perceptif (e.g. FFA, OFA; Elbich & Scherf, 2017; L. Huang et al., 2014; Jiahui et al., 2018) et à certains moments du traitement perceptif en réponse à des visages (e.g. N170 Herzmann et al., 2010; J. Towler & Eimer, 2012) semblent liées à cette habileté. Observer une différence d'activation cérébrale, par contre, ne renseigne que très peu sur les mécanismes sous-jacents. En effet, si ces études nous renseignent sur l'emplacement de cette information (le *où*), elles ne nous renseignent pas sur les mécanismes cérébraux et computationnels qui sont à l'œuvre dans la mise en place de ces différences (le *comment*). Le code cérébral des individus les plus habiles en reconnaissance faciale (i.e. *super-recognisers*; Russel, 2009) et des individus avec une perte significative de cette habileté (i.e. prospagnosiques; Bodamer, 1947) reste ainsi totalement inconnu. À ce niveau, la modélisation du cerveau avec l'apprentissage automatique (i.e. *machine learning*) et des modèles d'apprentissage profonds (i.e. *deep learning*) offre une approche robuste et sensible à des variations de computations dans les représentations cérébrales. Associer ces modèles computationels à l'imagerie cérébrale semble, ainsi, particulièrement approprié pour explorer ces mécanismes cérébraux interindividuels.

L'objectif de cette thèse est de mieux comprendre les mécanismes perceptifs et cérébraux qui supportent les différences individuelles en vision à l'aide d'outils psychophysiques, de l'imagerie cérébrale, et de l'apprentissage automatique. Celle-ci traitera spécifiquement des différences individuelles en reconnaissance faciale, un sujet fascinant tant pour sa complexité que

pour l'importance qu'il occupe dans notre quotidien. Nous aurons la chance de pouvoir caractériser ces variations sur des individus provenant de différentes populations diversifiées : des population cliniques (i.e. individus sur le spectre de la schizophrénie [article 1], un cas important de prosopagnosie acquise [article 5]), des individus neurotypiques [article 2,3,4], ainsi que des individus avec un système visuel exceptionnel (i.e. des « *super-recognisers* » dans le top 1% du spectre d'habileté [article 4]). Ces cinq articles toucheront tour à tour à plusieurs questions :

- Quel contenu visuel spécifique influence l'habileté d'individus typiques et de ceux avec des troubles psychiatriques?
- Le contenu visuel est-il plus ou moins modulé entre différentes tâches chez des individus performants?
- Peut-on prédire l'habileté en reconnaissance faciale d'un individu seulement à partir de quelques secondes d'activité cérébrale ?
- Au delà des représentations et du contenu visuel, quelle est la nature des computations cérébrales influençant l'habileté perceptive?
- Si notre cerveau effectue des computations similaires à celles de modèles d'apprentissage profond visuels, dans quelle mesure cette similarité est reliée à notre habileté en reconnaissance faciale?

À travers ces questions, nous nous attarderons ainsi particulièrement à trois types de mécanismes qui peuvent affecter ces différences individuelles : i) le *contenu représentationnel* ii) les *dynamiques cérébrales* et iii) les *computations cérébrales*. Les buts détaillés des articles qui contribueront à éclaircir ces mécanismes seront décrits dans la dernière section (1.5) de l'introduction. Avant, nous passerons à travers un survol des concepts et méthodes centraux à cette thèse : les différences individuelles en vision, le contenu représentationel, les représentations/dynamiques cérébrales, les techniques d'analyses multivariées, et les modèles d'apprentissage profonds visuels et langagiers.

1.2 Différences interindividuelles cérébrales et perceptives

Les différences individuelles en vision sont omniprésentes. Le cerveau de chaque individu contient des représentations distinctes de visages, mais aussi de différents objets familiers et de scènes qui lui sont propres. Ces variations interindividuelles émergent dès les premières étapes de notre perception, de la structure macroscopique du cortex visuel jusqu'aux répercussions de ces différences sur ses capacités : e.g. la surface corticale (Bergmann et al., 2016; Schwarzkopf et al., 2011), le contenu subjectif de la conscience visuelle (Stein et al., 2016), la mémoire perceptive et l'habileté de reconnaissance d'objets (Gauthier, 2018; Gauthier, Skudlarski, et al., 2000; Wilmer et al., 2012) varient tous énormément d'un individu à l'autre. La taille de la surface du cortex visuel primaire (V1), par exemple, varie d'un facteur de 3 entre individus (Schwarzkopf et al., 2011). Cette variation est prédictive de la taille subjective perçue (e.g. illusion de taille d'Ebbinghaus, voir **figure 1.1**), de même que la clarté et la précision à laquelle un individu crée une image mentale (Bergmann et al., 2016) (Schwarzkopf et al., 2011); (Bergmann et al., 2016). Toujours à titre d'exemple, la capacité de stockage en mémoire de travail visuelle varie avec l'âge (Brockmole & Logie, 2013), selon différentes populations cliniques (e.g. schizophrénie, voir (Forbes et al., 2009); Alzheimer; (Liang et al., 2016); revue : (Fukuda et al., 2015; Vogel & Machizawa, 2004), mais également au sein de la population jeune et saine (D. E. Anderson et al., 2011); (Fukuda et al., 2015; Vogel & Machizawa, 2004).

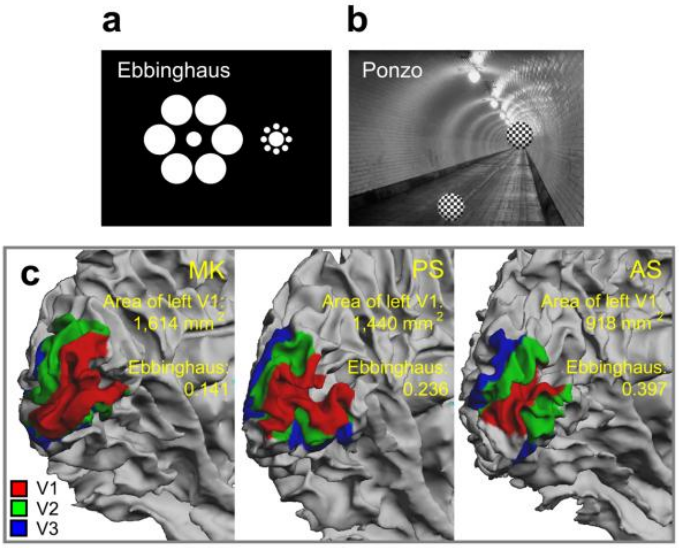


Figure 1.1 Variations individuelles du cortex visuel primaire associé à la taille subjective perçu. Des variations individuelles dans la taille de la surface corticale du cortex strié (c) sont prédictives de la taille subjective perçue (a,b). Plus la surface de V1 est petite chez un individu, plus il aura tendance à percevoir le percept illusoire de différence de taille. Figure provenant de (Schwarzkopf et al., 2011).

Ces différences individuelles, par contre, n'ont été que très récemment étudiées par les chercheurs en neurosciences cognitives (Dubois & Adolphs, 2016; Wilmer et al., 2012). Elles ont en effet traditionnellement été traitées comme une source de bruit dans différents contextes expérimentaux. Ainsi, les variations dans des processus aussi fondamentaux que ceux qui nous permettent de reconnaître les visages et les objets de notre entourage, même s'ils sont très variables d'un individu à l'autre (Charest & Kriegeskorte, 2015; Sigurdardottir & Gauthier, 2015; Wilmer et al., 2012), restent largement inconnus. Parmi ces variations, celles dans l'habileté de reconnaissance faciale offrent un exemple frappant d'un processus sur lequel tout individu repose mais qui varie de façon importante entre individus. Certains individus, appelés *super-recognizers* (Russell et al., 2009), ont une habileté extraordinaire en reconnaissance faciale qui leur permettent de mémoriser l'identité faciale d'étrangers aperçus seulement une fois. Pour d'autres, à l'inverse, cette habileté est très limitée (*prosopagnosie(McConachie, 1976; Susilo & Duchaine, 2013)*. Ces individus se placent sur un spectre d'habileté perceptive en reconnaissance faciale, tel que celui présenté dans la **figure 1.2**, où les super-recognisers se positionnent dans le top 2% de la distribution (Russell et al., 2009) et les prosopagnosiques en dessous du 2ème percentile d'habileté. De telles différences ont évidemment

un impact important au niveau des habiletés sociales, et du bien être des individus (Dalrymple et al., 2014). La prochaine section introduira certains des mécanismes perceptifs derrière l'habileté en reconnaissance faciale.

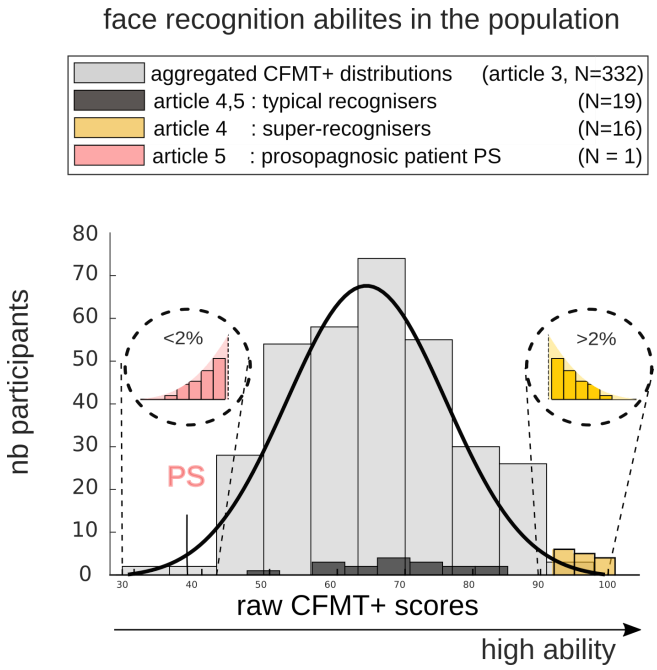


Figure 1.2. Distribution de l'habileté individuelle en reconnaissance faciale. Les scores individuels au *Cambridge Face Memory test long-form* (CFMT+, (Russell et al., 2009) compilés à travers différentes études (Faghel-Soubeyrand, Alink, Bamps, Gervais, et al., 2019; Fysh et al., 2020; Tardif et al., 2019a), incluant le troisième article de cette thèse, montrent les variations individuelles d'habileté en reconnaissance faciale. Les échantillons des différents articles incluent dans cette thèse sont également représenté sur cette distribution, de la prosopagnosie (rouge; voir article 5) jusqu'aux super-recognisers (jaune; voir article 4).

1.2.1 Mécanismes perceptifs des variations individuelles en reconnaissance faciale

L'habileté en reconnaissance faciale est un trait stable, héritable, et relativement indépendante de l'intelligence générale d'un individu (Wilmer et al., 2010, 2012). Plusieurs dimensions cognitives et perceptives ont été tentativement associées à ces différences. Notamment, le concept populaire de traitement visuel « holistique » qui repose sur l'idée que l'intégration d'attributs visuel est plus que la somme de ses parties (Maurer et al., 2002; Richler & Gauthier, 2014; Tanaka & Gordon, 2011). Ces tentatives, par contre, n'ont révélés que des corrélations soit très faibles (DeGutis et al., 2013; Wang et al., 2012), ou inexistantes (Konar et al., 2010) avec l'habileté en reconnaissance faciale (possiblement à cause de difficultés dans la façon d'opérationnaliser ce concept ou des nombreuses définitions qui ont été utilisées au court des années; voir Richler et al., 2015). D'autres pistes ont été plus fructueuses. Plusieurs évidences, notamment certaines provenant du mémoire de M.Sc. du présent candidat (Faghel-Soubeyrand, Dupuis-Roy, et al., 2019), indiquent que des variations individuelles dans le *contenu visuel* de nos représentation des visages influencent cette habileté. Ces évidences seront décrites dans la prochaine section. Nous traiterons spécifiquement plus en profondeur d'une famille de techniques expérimentales permettant d'extraire ce contenu représentationel, la *reverse correlation.* Nous verrons que cette relation entre contenu représentationel et habileté semble tenir à travers des populations diversifiées, notamment chez certaines populations avec des troubles fonctionnels (psychiatriques) significatifs.

1.2.2 Contenu représentationnel visuel et différences individuelles

Imaginez une scène visuelle d'une plage ensoleillée avec des parasols, des vacanciers, du sable blanc, et la mer à perte de vue à l'horizon. Si on vous demande de trouver « Charlie » au sein de cette scène, vous devrez sans doute isoler de l'information spécifique (e.g. la couleur, puis les formes distinctes d'individus, etc.). Il est possible de révéler l'information visuelle physique qu'un

individu extrait pendant une telle tâche. Nous référons à cette information visuelle liée à une représentation mentale spécifique en tant que contenu représentationnel (Schyns et al. 2020). Nous savons que le contenu représentationel est lié à l'habileté de reconnaissance dans la population. Des variations dans l'utilisation de la structure horizontale, de même que dans l'utilisation de la région des yeux d'un visage, par exemple, sont corrélées à de meilleures habiletés en reconnaissance faciale (Faghel-Soubeyrand, Dupuis-Roy, et al., 2019; Pachai et al., 2013; Royer et al., 2018; Tardif et al., 2019b). Nous avons même pu démontrer que de forcer l'utilisation d'un contenu représentationel spécifique à des individus *cause* de meilleures performances en reconnaissances faciales (Adolphs et al., 2005; Faghel-Soubeyrand, Dupuis-Roy, et al., 2019). Une grande partie de ces avancées ont été faites à l'aide d'une technique de *reverse correlation* permettant de faire le pont entre l'information physique et la catégorisation: la technique *Bubbles* (Gosselin & Schyns, 2001).

Cette technique a également permi de caractériser les représentations visuelles d'individus avec des troubles fonctionnels significatifs. En utilisant cette technique, Adolphs et coll. (2005) ont par exemple mesuré l'utilisation d'information pendant une tâche d'identification émotionnelle chez la patiente SM, une femme cérébro-lésée aux deux amygdales (une région cérébrale classiquement associée au traitement de l'émotion de la peur). Comparée aux individus sains qui utilisent significativement la région des yeux pour identifier l'émotion de peur, SM utilise un contenu représentationnel sous-optimal (la région de la bouche) pour effectuer la même tâche. Cette différence de stratégie était accompagnée d'une performance en reconnaissance des expressions faciales significativement plus faible chez SM comparée aux contrôles. Quand les chercheurs ont explicitement induit l'utilisation des yeux chez SM, sa performance devenait similaire à la moyenne. Des résultats similaires ont été observés chez un des cas les plus spécifique et documenté de *prosopagnosie acquise* (Bodamer, 1947), la patiente PS (Rossion, 2022a, 2022b; Rossion et al., 2003). Une description détaillée de la patiente PS, une femme avec des déficits virtuellement complet de reconnaissance faciale d'identité, est présenté dans l'article 5. Suite à des lésions cérébrales dans le système perceptif (gyrus fusiforme [FFA] gauche, [OFA] droit), PS n'arrivait plus à reconnaître son propre visage dans le miroir ou celui de ses proches. De façon intéressante, en

utilisant la même méthode des bulles, Caldara et al., (2005) ont montré que le contenu des représentations visuelles de PS sous-représentait les yeux pour reconnaître l'identité d'un visage. De telles différences dans l'utilisation d'information ont aussi été révélées chez des populations cliniques telles que chez des individus sur le spectre de la schizophrénie (Clark et al., 2013; Lee et al., 2011) et de l'autisme (Spezio et al., 2007; Alink & Charest, 2020), ainsi que chez des individus atteints d'anxiété sociale (Langner et al., 2009). L'article 1 de cette thèse (Faghel-Soubeyrand et al., 2020) révèle pour la première fois le contenu représentationnel et les déficits perceptifs d'individus avec des troubles fonctionels important ayant un diagnostic comorbide de schizophrénie *et* d'anxiété sociale.

L'ensemble de ces études (incluant l'article 1 de la présente thèse) indique ainsi que l'information spécifique qui est contenue dans une représentation visuelle—le contenu représentationnel—diffère de façon importante entre individus, et que de telles variations du contenu représentationnel sont reliées à des processus fondamentaux de la perception, de la structure cérébrale et du fonctionnement général d'un individu. Deux points manquent par contre à l'ensemble de ces études.

Premièrement, les résultats ayant associé des variations dans l'habileté perceptive au contenu représentationnel n'ont typiquement utilisé qu'une seule dimension de paramètre du contenu (e.g. l'information d'orientation *ou* de position dans Pachai et al., 2013; Duncan et al., 2017). Ceci rend non seulement impossible de quantifier le degré d'implication de différents contenus représentationnels dans ces variations (e.g. information spatiale *vs.* orientation), mais nous empêche également de comprendre les interactions entre divers types de contenu (e.g. information spatiale *et* orientation). Il est probable que les résultats montrant que le contenu représentationnel spatial soit lié à des différences d'habileté (c.-à-d. les yeux, dans Faghel-Soubeyrand, Dupuis-Roy, et al., 2019; Royer et al., 2018; Tardif et al., 2019b) et ceux montrant une telle association avec le contenu d'orientation (Duncan et al., 2017; Pachai et al., 2013) soient l'envers d'une même médaille. Au niveau cognitif, des études récentes suggèrent que l'information d'orientation horizontale, par exemple, est particulièrement importante pour le traitement facial (e.g. voir Goffaux & Greenwood,

2016; Jacques et al., 2014; Pachai et al., 2013; Taubert et al., 2016). Afin de mieux comprendre les variations du contenu représentationnel à travers la population, l'article 2 de cette thèse a mesuré simultanément trois dimensions physiques du contenu visuel: l'information d'orientation, de positions et de fréquences spatiales (FS). Ces trois caractéristiques sont centrales au régions cérébrales du système visuel (e.g. Hubel & Wiesel, 1968; Jones & Palmer, 1987). Comme nous le verrons dans l'article 2 (Faghel-Soubeyrand et al., 2021), une telle caractérisation des trois dimensions visuelles n'a jamais été faite simultanément pendant une tâche de reconnaissance faciale, et nous donne un portrait complet du contenu représentationnel pendant une tâche de reconnaissance faciale.

Deuxièmement, l'ensemble de ces études a pu démontrer qu'*au sein même d'une tâche de reconnaissance* (e.g. reconnaissance d'expression faciale dans Faghel-Soubeyrand et al., 2020), l'habileté en reconnaissance faciale est corrélée avec un contenu facial spécifique. Mais cette approche omet un point important du monde visuel; les demandes du monde externe sont typiquement multiples, et le cerveau extrait de l'information à la fois sur l'identité d'un visage, mais également de l'information sur le genre, l'expression faciale, ou la confiance d'un visage présenté, par exemple. Ainsi, en ce moment, nous n'avons aucune indication de la stratégie visuelle qui est effectuée par un même individu habile *à travers différentes tâches en reconnaissance faciale.* L'article 3 testera l'hypothèse que les meilleurs participants utilisent des représentations visuelles plus différentes à travers les tâches. Nous avons mesuré ces différences de représentations visuelles entre tâches — reconnaissance i) d'expressions et ii) du genre d'un visage — dans un large échantillon neurotypique (N=120) avec la méthode psychophysique validée dans l'article 2. Nous pourrons ainsi, d'une part, caractériser le contenu représentationnel d'un large échantillon, mais également associer des variations du contenu représentationnel à l'habileté du système visuel d'un individu.

1.3 Représentations et dynamiques cérébrales

S’ils nous renseignent sur le contenu des représentations visuelles reliées à l’habileté, les études et méthodes décrites plus haut ne permettent pas de révéler les mécanismes cérébraux qui supportent ces variations. Nous savons pourtant que le traitement d’une image implique une série de nombreuses transformations dès la première salve de traitements ascendants la hiérarchie visuelle (voir figure 4 ci-dessous): des cellules ganglionnaires de la rétine au thalamus (CGL), aux aires postérieures du système visuel (e.g. cortex strié, V1), puis aux aires antérieures de la voie ventrale (e.g. cortex inféro temporal humain, hIT; Cichy et al., 2014; Tanaka, 1993). Une des premières transformations importantes d’une image suite à son passage par la rétine est effectuée par V1. Cette transformation est similaire à une analyse de Fourier discrète: les caractéristiques de réponses des différentes populations de neurones dans V1 (Hubel & Wiesel, 1962, 1968; Kay & Yeatman, 2017; Roth et al., 2022) résultent en une série de réponses ayant des fonctions similaires à des convolutions avec des filtres sinusoïdales de différentes tailles, orientations, fréquences spatiales et positions préférentielles. Cette décomposition d’une image en une somme finie de grilles sinusoidales, centrale en neuroscience visuelle et considérée comme une information cruciale pour le reste du système visuel, est utilisée dans les techniques de *reverse correlation* décrites dans les articles 1, 2 et 3 pour échantillonner le contenu représentationnel (Gosselin and Schyns 2001; Faghel-Soubeyrand et al. 2021; Faghel-Soubeyrand et al. 2020; Alink and Charest 2020). On dira que cette transformation d’une image par V1 est une transformation de “*bas-niveau*” (d'abstraction) parce qu’elle est basée sur des propriétées physiques simples de l’image (e.g. la position rétinotopique; Groen et al., 2017).

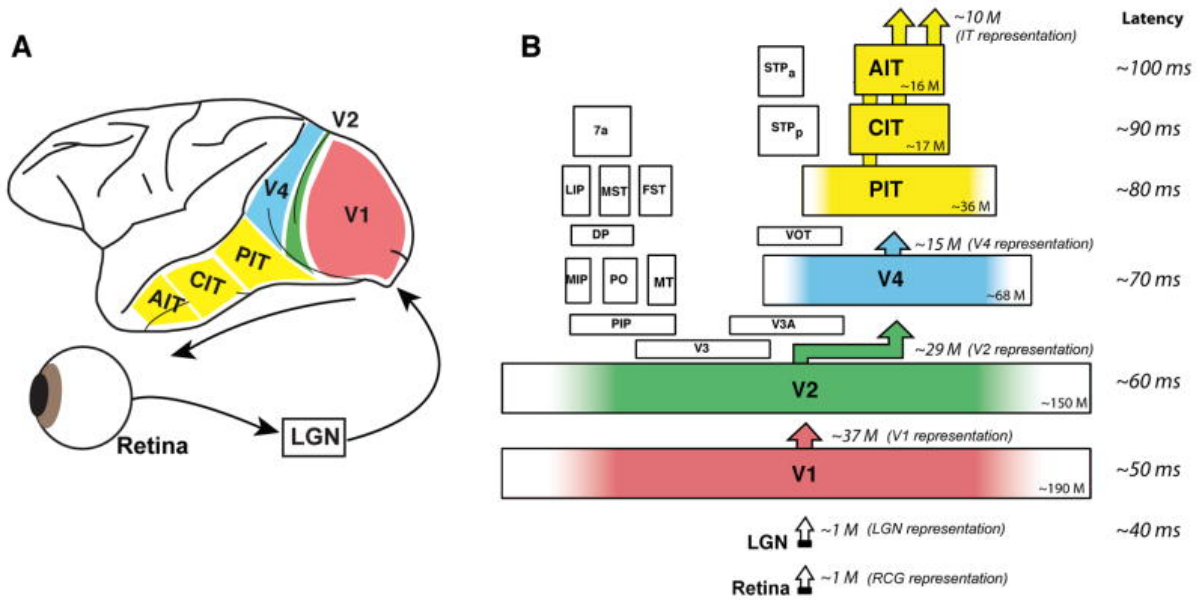


Figure 1.3. Traitement visuel de la rétine aux aires antérieurs du système visuel ventral. a) Le schéma décrit le flot ascendant de l'information visuelle, dans le cortex du système visuel ventral du macaque, suite à sa réception sur la rétine. **b)** Les latences de l'activité neuronales dans les différentes régions cérébrales du système visuel suite à la présentation d'une image sont indiquées sur la droite. Figure provenant de DiCarlo et al., (2012).

Plus en aval dans le traitement visuel, les neurones des régions d'hIT répondent quant à eux davantage à des stimuli de catégories spécifiques, ou à tout le moins encodent une image selon des computations plus abstraites où la dimensionnalité de la réponse est fortement réduite (DiCarlo et al., 2012; Hebart et al., 2018; Jozwik et al., 2016; Rust & DiCarlo, 2012; Tanaka, 1993). Les études pionnières en imagerie cérébrale fonctionelle (IRMf/*Position Emission Tomography*, PET) ont en effet révélé différents modules spécialisés pour le traitement de stimuli spécifiques au sein de hIT: le *Fusiform Face Area*, une région sensible aux visages (FFA, Kanwisher et al., 1997); le *Parahippocampal Place Area*, une région sensible aux scènes (PPA; Epstein et al., 1999); et le *Lateral Occipital Complex*, une région sensible aux objets (LOC, Grill-Spector et al., 2001; Malach et al., 1995). La conception populaire en neuroscience visuelle est ainsi que c'est dans ces régions antérieures de hIT où s'effectue la reconnaissance faciale (FFA), de scènes (PPA), ou d'objets (LOC). Notons que c'est notamment dans hIT qu'on retrouve des régions répondant significativement à l'*identité* d'un visage, peu importe sa position, son orientation, et sa taille (Tsao

& Livingstone, 2008). On considérera alors que la réponse dans ces régions sera de plus *haut-niveau* (d'abstraction) parce qu'elles sont sensibles à des attributs visuels plus complexes et relativement invariants à des attributs visuels plus simples (e.g. visages dans le FFA vs. position, orientation, et FS dans V1). Le traitement d'un stimulus visuel par le cerveau implique ainsi une chaîne de processus qui lie l'information visuelle riche reçue sur la rétine puis traitée par les régions postérieures, à un code de plus en plus abstrait et invariant en passant à travers les régions antérieures (M. L. Smith et al., 2012; Ullman et al., 2002). Cette idée de traitement hiérarchique des représentations cérébrales, même si elle devrait être considérée avec nuance, est centrale aux neurosciences-cognitives visuelles (elle semble également émerger naturellement chez des modèles artificiels de reconnaissance d'objets; voir section 1.4). Ces transformations de plus en plus abstraites s'effectuent évidemment aussi à travers le temps dans le cerveau. Des études utilisant des méthodes invasives chez le singe (Bullier, 2001; DiCarlo et al., 2012; Schmolesky et al., 1998) ont montré que suite à la présentation d'une image, des neurones dans V1 s'activent avec un délai de ~50 ms, suivit de V2 ~60 ms , IT ~80ms, et des régions antérieurs de IT vers 100 ms (voir figure 1.3). Chez l'humain, l'électroencéphalographie (EEG) et la technique des potentiels évoqués (Event Related Potentials, ERP; (Boudewyn et al., 2018; Luck et al., 1990; Rossion & Jacques, 2012; M. L. Smith et al., 2012; M. L. Smith, Gosselin, & Schyns, 2004) sont certainement les méthode les plus utilisées pour capturer cette hiérarchie temporelle (voir également les processus oscillatoires reliées à ces dynamiques; (Buzsáki, 1996). Ces techniques ont révélé des fenêtres temporelle correspondant à un traitement cérébral précoce, intermédiaire, et tardif, i.e. la P100 (Luck et al., 1990), N170 (Bentin et al., 1996) and N400 components (Kutas & Federmeier, 2000).

1.3.1 Activité cérébrale univariée

Les résultats ayant établi ces modules cérébraux spécialisés pour des catégories d'objets, de même que ceux utilisant les ERPs, reposent tous sur une méthode d'analyse simple qui consiste à moyenner l'activité cérébrale (e.g. des *voxels* au sein de régions d'intérêt*s* en fMRI) afin d'observer les différences d'activation entre une catégorie de stimulus (e.g. visage) et un autre (e.g. des bâtiments). Ces techniques d'*analyses univariées* et d'*analyses de contrastes* ont été

particulièrement utiles dans un contexte où nous devions réduire le plus possible le bruit provenant de multiples sources physiologiques (e.g. différences de vascularisation pour le fMRI) et techniques (e.g. provenant du bruit électrique ambiant pour l'EEG). Ces techniques ont été, et restent toujours, la norme dans la majorité des études en neurosciences-cognitives et en psychologie — incluant la littérature sur la perception faciale. À l'heure actuelle, par exemple, le peu d'études portant sur les bases cérébrales de l'habileté visuelle en reconnaissance faciale ont utilisés des méthodologies qui ne précisaient qu'en surface les différences de traitement cérébraux entre individus. À titre d'exemple, nous savons que la sélectivité corticale aux visages dans le FFA (i.e. le *contraste* de la réponse IRMf aux visages vs. aux non-visages) corrèle avec l'habileté en reconnaissance faciale chez des participants neurotypiques (Huang et al., 2014). Cependant, la sélectivité aux visage dans le FFA n'explique pas l'habileté réduite des individus chez qui la prosopagnosie est acquise (e.g. voir résultats chez PS (Gao et al., 2019; Rossion, 2022b), ou même de prosopagnosiques développementaux (i.e. individus sans lésions cérébrales, e.g. Avidan et al., 2005; Behrmann et al., 2007; Minnebusch et al., 2009; Williams et al., 2007); mais voir Furl et al., 2011). Ces contradictions indiquent que de telles mesures sont pour le moins incomplètes ou trop peu sensibles pour rendre compte des processus complexes qui supportent les différences en reconnaissance faciale. Des résultats un peu plus convaincants ont été trouvés au niveau du traitement temporel dans le cerveau. Certaines associations (e.g. Alonso Prieto et al., 2011; Bobes et al., 2003; Herzmann et al., 2010; Kaltwasser et al., 2014; Towler & Eimer, 2012) ont été faites avec la composante ERP N170, une déflection de voltage négative dans les électrodes occipitotemporales entre 130 et 230 ms après la présentation d'un visage/objet (Bentin et al., 1996). Ces différences *suggèrent indirectement* que des mécanismes de traitement de niveaux intermédiaires sont liés à l'habileté. D'autres études ont, par contre, révélé une absence de différences avec des individus prosopagnosiques développementaux (e.g. Eimer et al., 2012). Crucialement, de façon générale, l'ensemble de ces études ayant utilisé des techniques univariées et de contrastes, que ce soit en EEG ou en fMRI, n'ont pas révélé les computations qui supportent ces différences d'activations.

Ces approches univariées et de « *localisers* » limitent en effet significativement une description riche et profonde des processus fonctionnels du cerveau humain en soustrayant de

l'activité cérébrale importante (Friston et al., 2006; Vinken et al., 2022). Un bon exemple de ces limites (Vinken et al., 2022) a récemment été montré avec les « *face-cells* » (Tsao et al., 2006), l'archétype des neurones sélectives à une catégorie dans le cortex inféro-temporal du singe. Nous savons que ces *face-cells,* même si elles sont activées significativement plus pour les visages que les non-visages, répondent quand même très fortement à la plupart des catégories d'images. Une étude récente de Vinken et al., (2022) a montré que la caratéristique même de sélectivité aux visage des *face-cells* est en fait mieux expliquée par la réponse aux *non-visages* que la réponse de ces neurones aux visages. Des analyses plus approfondies ont montré que des caractéristiques qui sont générales à tout type d'image (i.e. de l'information domaine-générale similaire à celle de réseaux d'apprentissage profond entraîné sur la reconnaissance d'objets; voir aussi Long et al., 2018) explique cette relation. Il semble ainsi que des approches investiguant des processus cérébraux avec des designs expérimentaux riches en conditions (Allen, et al., 2022; Charest et al., 2014; Kriegeskorte & Kievit, 2013; Naselaris et al., 2021) et qui rendent compte d'une description plus large des mécanismes cérébraux sous-jacents en testant divers modèles sur l'ensemble du cerveau (Dwivedi et al., 2021 ; Kriegeskorte & Diedrichsen, 2019; Popham et al., 2021) permettent une compréhension beaucoup plus complète de ces processus perceptifs complexes.

1.3.2 Représentations cérébrales et analyses multivariées

L'avènement d'analyses de patrons multivariées en neuroimagerie (Haxby et al., 2001, 2014) a permis de dépasser l'association simple entre régions/stimuli en répondant à des questions référant au type et à l'organisation du code dans une région donnée (Haxby et al., 2001, 2014). Diverses études utilisant la MVPA ont ainsi montré que des différences d'activations subtiles dans V1 – indétectables par analyses univariées – contiennent de l'information permettant de discriminer des caractéristiques précises de bas niveaux, telles que l'orientation (Kamitani & Tong, 2005), de l'information visuelle gardée en mémoire de travail (Harrison & Tong, 2009), des représentations imaginées (i.e. l'imagerie visuelle; Naselaris et al., 2015; Reddy et al., 2010), des mélodies (Saari et al., 2018), les niveaux de consciences (Levinson et al., 2021), l'émotion ressentie (Hofmann et al., 2021; Kragel & LaBar, 2016) et beaucoup d'autres (Haxby et al., 2014). Plusieurs familles d'analyses multivariées sont pertinentes à l'étude des représentations cérébrales en neurosciences

cognitives et computationelles (Kriegeskorte, 2015b; Kriegeskorte & Diedrichsen, 2019; Naselaris et al., 2011; Paninski et al., 2007). Parmi celles-ci, deux sont tout particulièrement d'intérêt pour cette thèse : le *decoding,* et la *Representational Similarity Analysis* (RSA). Les deux prochaines sous-sections auront pour but non pas de développer sur ces techniques — des revues de littératures détaillées ont déjà été publiées sur ces méthodes (Haxby et al., 2014; Kriegeskorte & Diedrichsen, 2019) — mais plutôt de préparer le lecteur au contenu des articles de cette thèse.

1.3.2.1 Lire l'information cérébrale : le decoding

Une des premières questions que notre équipe s'est posée est la suivante : peut-on déterminer l'habileté en reconnaissance faciale d'un individu seulement à partir de son cerveau ? Cette question peut paraître simple. Y répondre avancerait par contre drastiquement notre compréhension des différences individuelles en perception, pourrait améliorer la qualité de vie d'individus prospagnosiques en permettant un diagnostic objectif rapide, et pourrait même faciliter le repérage d'expert.e.s en reconnaissance faciale pour certains emplois en sécurité (e.g. forces de police, service de douanes et de sécurité).

Un modèle de *decoding* (aussi appelé *decoder* ou *classifier*) permet de prédire une caractéristique d'intérêt à partir d'un patron d'activité cérébrale (Grootswagers et al., 2017; Kriegeskorte & Diedrichsen, 2019). Tant l'*output* prédictif du modèle, typiquement une condition expérimentale (e.g. une image perçue; Hung et al., 2005; Kay et al., 2008), que son *input* (ici des patrons d'activité cérébrale) peuvent être extrêmement variés dans leur format. Ce format/espace représentationel est variable entre les méthodes d'imageries qui enregistrent les patrons d'activité neuronales; les attributs ou *features* en EEG sont des patrons d'activités provenant d'électrodes à différents moments dans le temps (e.g. 64 électrodes x 100 échantillons, donc 6400 *features* possibles). Ceux en IRMf proviennent typiquement de centaines ou de milliers de *voxels* dans une région cérébrale donnée (e.g. le FFA dans Zhang et al., 2015). On peut ainsi tenter de prédire des caractéristiques simples, telles que l'âge ou le sexe, à travers un échantillon de participants à partir de leur signal d'IRMf, d'enregistrement intracrâniens (Proix et al., 2022), ou de signal électroencéphalographique (EEG; Grootswagers et al., 2017; Guggenmos et al., 2018), dépendament

des besoins expérimentaux. L'idée sous-jacente cruciale dans le contexte de cette thèse est que si nous sommes capables de décoder notre caractéristique d'intérêt — l'habileté perceptive — à partir de signal cérébral de participants en réponse à un stimulus visuel donné (e.g. une scène ou un visage dans la figure 1.4b), le cerveau de ces individus (i.e. le traitement du cerveau à un moment donné pour l'EEG dans la figure 1.4) contient de l'information sur cette caractéristique.

Parmi les différents types d'algorithmes de classification automatiques, les décodeurs linéaires sont sans doute les plus utilisés avec les données de neuroimagerie. Au-delà de leur simplicité et leur efficacité d'implémentation algorithmique comparé aux modèles non-linéaires, ils reposent sur la caractéristique intuitive d'une lecture linéaire de l'activité cérébrale (*linear read-out*; Kriegeskorte & Diedrichsen, 2019). En effet, on dira que l'output d'une population de neurones (e.g. V1) est « interprétable » par la région en aval du traitement (e.g. hIT) quand elle peut être séparée linéairement par celle-ci (DiCarlo et al., 2012). La création de ces modèles de décodeurs linéaires repose sur une série d'opérations généralement assez simples ayant pour but trouver le meilleur fit entre un set d'observations $O_{i...n}$ et les caractéristique associées $F_{i....n}$ (e.g. l'habileté perceptive) à travers un modèle de poids P (e.g. une matrice de dimension égale au nombre de canaux). Chaque observation O_i , une fois le modèle P produit, n'aura qu'à être projetée sur celui-ci pour retrouver la valeur F prédite F'($Fi' = P(Oi) + b$, où b est un terme de biais). La figure 1.4 illustre ce problème avec notre question initiale, traitée empiriquement dans l'article 4. Nous désirons retrouver l'habileté perceptive (F) à partir d'un ensemble d'observations $O_{i...n}$ — des milliers de patrons d'activité EEG de 128 canaux à un moment donné provenant de *super-recognisers* (points jaunes) et d'individus typiques (points gris). Des décodeurs *linear discriminants* sont utilisés, à chaque moment, pour créer un espace décisionnel P qui optimise la distance entre les deux classes d'individus (Treder, 2020). Si elle opère avec succès, cette opération nous donnera une performance au-delà de la chance (ici 50% pour deux classes), et nous permettera ainsi de *décoder l'habileté perceptive à partir du cerveau d'un individu.*

Cette technique d'analyse a connu un grand succès depuis une vingtaine d'années. Un article percutant ayant utilisé cette technique en combinaison avec des données d'IRMf (Horikawa et al., 2013) a même été capable de lire le contenu sémantique des rêves d'individus pendant leur sommeil

(voir aussi la détection de rêve en EEG haute-densité, Siclari et al., 2017). C'est pourquoi on réfère parfois aux modèles de decoding comme du "*mind reading*"; on peut lire, à travers l'activité cérébrale, le contenu de la perception et de plusieurs autres états subjectifs (Haxby et al., 2014; Haynes & Rees, 2006). Peu d'études, par contre, se sont attardées à prédire les différences interindividuelles à partir du cerveau (Banville et al., 2021), et aucune études n'a, à ce jour, décodé l'habileté perceptive individuelle (Zhang et al., 2015). Qui plus est, ce défi devient potentiellement encore plus considérable si on veut y répondre avec une méthode de neuroimagerie accessible et possédant une bonne résolution temporelle (i.e. à relativement faible coût, non-invasive, et rapide) telle que l'EEG. En effet, le *decoding* cérébral est typiquement effectué sur des données de relativement haute résolution spatiale (e.g. avec plusieurs centaines de voxels en searchlight d'IRMf; Haxby et al., 2014; Zhang et al., 2015). Par contre, les dynamiques cérébrales — dimension temporelle constituant un aspect fondamental des processus cérébraux — sont virtuellement non mesurables par ces techniques d'imagerie. La caractérisation des dynamiques cérébrales avec ces décodeurs rajouterait à notre compréhension des processus dans la hiérarchie de traitement cérébral énoncé plus tôt. Comme nous le verrons dans les articles 4 et 5 de cette thèse, être capable de décoder l'habileté perceptive avec l'EEG apporterait ainsi plusieurs avancées au niveau théorique, méthodologique, et même potentiellement clinique en perception visuelle (voir la section Discussion générale de cette thèse).

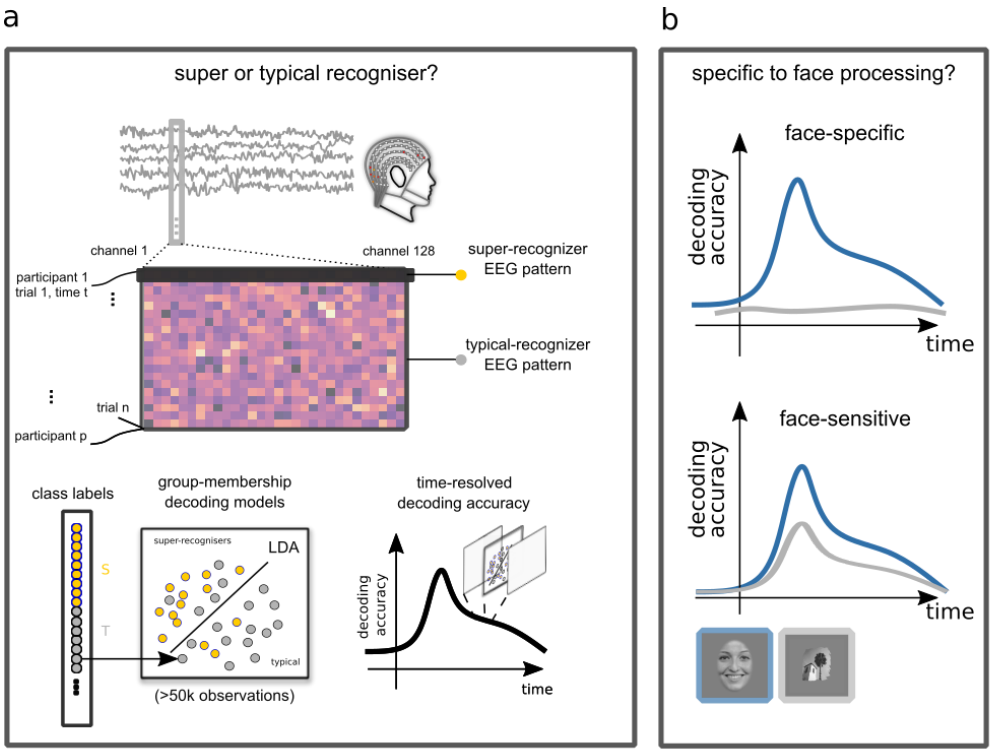


Figure 1.4. Illustration du procéder permettant de décoder l'habileté perceptive à partir de l'EEG. (a) Des patrons d'activité EEG bruts en réponse à différentes images (e.g. un visage) sont compilés à travers les participants. Chaque patron d'activité cérébral est associé à une classe; dans l'article 4, le groupe de *super-recognisers* (jaune) ou le groupe avec une habileté en reconnaissance faciale typique (gris). Un algorithme de linear discriminant trouve la projection qui maximise la distance entre les observations des deux classes (>50,000 dans l'article 4), tout en minisant la distance intra-classe. Répéter ce processus à travers les pas de temps échantillonnés en EEG nous permet d'observer le décours temporels des représentations cérébrales associées à l'habileté perceptive.

1.3.2.2 Révéler la structure générale des représentations cérébrales : le code représentationnel

D'autres méthodes d'analyses multivariées rendent honneur aux idiosyncrasies du cerveau et du système visuel (Charest et al., 2014). Une famille particulière d'analyses multivariées, appelée *Representational Similarity Analysis* (RSA, Kriegeskorte et al., 2008; Kriegeskorte & Kievit, 2013), permet de révéler comment une région représente l'information lui permettant de traiter différentes conditions expérimentales. L'idée générale est simple : en révélant la (dis)similarité entre des patrons d'activités cérébraux provenant d'une multitude de différentes conditions (typiquement des stimuli), nous pouvons comprendre comment le cerveau réussi à les encoder. Pour l'EEG, par

exemple (voir figure 1.5a), on produira la dissimilarité des patrons d'activations topographiques [O_{Si} vs. O_{Sj}] en réponse à deux stimulus, S*i* et S*j*. En produisant la dissimilarité pour toutes les paires de stimuli de l'espace expérimental, on créera une Matrice de Dissimilarité Représentationelle (*Representational Dissimilarity Matrix*, RDM); une matrice symétrique ayant comme taille le nombre de stimuli sur ses rangées et ses colonnes (voir figure 1.5a). Parce qu'on peut considérer chacune des valeurs de dissimilarité comme des *distances représentationelles*[1] (e.g. la représentation cérébrale multivariée d'un visage peut être très distante de celle d'une maison dans hIT), l'ensemble de ces distances constitue une organisation des représentations d'un modèle donné qu'on appelle la *géométrie représentationelle*, ou le *code représentationnel.* Un exemple de géométrie représentationnelle est montré à la figure 1.5 pour des données en EEG de l'article 4. Cette analyse permet de décrire de décrire, par exemple, la capacité du cerveau de représenter différents stimuli de différentes catégories à travers le temps, et de révéler des distinctions importantes comme celle entre les visages et des non-visages (figure 1.5b, une caractéristique clef de la composante N170, Rossion & Jacques, 2012).

L'utilisation de la RSA avec des données d'IRMf a démontré comment certaines catégories functionelles (e.g. la distinction animée vs. inanimé, des visages vs. objets) émergent naturellement de l'information multivariée contenue dans le cortex inféro temporal (Cichy et al., 2014, 2019; Contini et al., 2017). Plus généralement, la RSA a ainsi pu mener à une meilleure compréhension de processus tels que la mémorabilité d'objets (Khaligh-Razavi et al., 2016), la familiarité des visages (Dobs et al., 2019) et l'émergence d'information basique vers l'information plus abstraite au sein du cortex (Carlson et al., 2019; Cichy et al., 2014; Contini et al., 2017; Hebart et al., 2018; voir Contini et al., 2017 pour une revue).

Cette technique peut être considérée comme une généralisation des techniques de *decoding* parce que sa caractérisation des combinaisons de paires de stimulus renseigne sur la structure générale à travers tous les exemplaires de toutes les catégories (Kriegeskorte & Wei, 2021). La RSA rend ainsi

[1] Notons que quand des distances sont mesurées à partir de données cérébrales (nécessairement bruitées), le bruit cause un biais qui exagère artificiellement cette distance; deux patrons d'activités bruités, par exemple, auront une corrélation resque nulle et une grande distance représentationelle. Pour éviter la confusion entre distance représentationelle et bruit, l'ensemble des mesures de distances représentationelles utilisées dans cette thèse proviendront de distances *cross-validées (Kriegeskorte & Diedrichsen, 2019)*.

compte du caractère multivarié des représentations cérébrales (comme le *decoding*), mais également de la richesse des conditions expérimentales (e.g. plus de 70,000 stimuli pour le Natural Scene Dataset; Allen et al., 2022).

Deux avantages de la RSA se distinguent toutefois par rapport à l'objectif de cette thèse. Premièrement, cette technique semble capturer le caractère unique des représentations cérébrales idiosyncratiques en vision (Charest et al., 2014; Charest & Kriegeskorte, 2015) ainsi que pour des processus de plus haut niveaux (Anderson et al., 2020; Popal et al., 2019). Une étude particulièrement intéressante de (Lee & Geng, 2017) a étudié le lien entre les différences individuelles en géométrie représentationnelles en fMRI et la performance dans l'allocation d'attention pendant une tâche de recherche d'identité faciale. Malgré le faible échantillon (N=14) et le nombre réduit de stimuli, ils ont pu montrer que des variations individuelles dans la géométrie représentationelle du cortex préfrontal dorso-latéral (dl-PFC) prédisaient la performance de discrimination entre deux identités faciales.

Deuxièmement, et c'est certainement l'avantage le plus important de cette technique, *la RSA permet de comparer des représentations ayant des formats totalement différents* (Cichy et al., 2014; Kriegeskorte et al., 2008; Kriegeskorte & Diedrichsen, 2016). En créant un format représentationnel — la RDM — qui s'affranchit du format représentationnel des données brutes, la RSA nous permet d'associer directement des représentations provenant de n'importe quelles modalités; différentes techniques d'imageries cérébrales (i.e. relié fMRI et EEG dans Cichy et al., 2014; Cichy & Oliva, 2020, différentes modalitées perceptuelles (e.g. verbales et d'attributs visuels dans Anderson et al., 2020). Cette capacité de la RSA de faire des ponts entre différentes modalités de représentations est centrale à cette thèse, puisqu'elle nous permettera d'associer les représentations cérébrales individuelles aux représentations hiérarchiques de réseaux de neurones artificiels de la reconnaissance d'objets et du traitement langagier (articles 4 et 5).

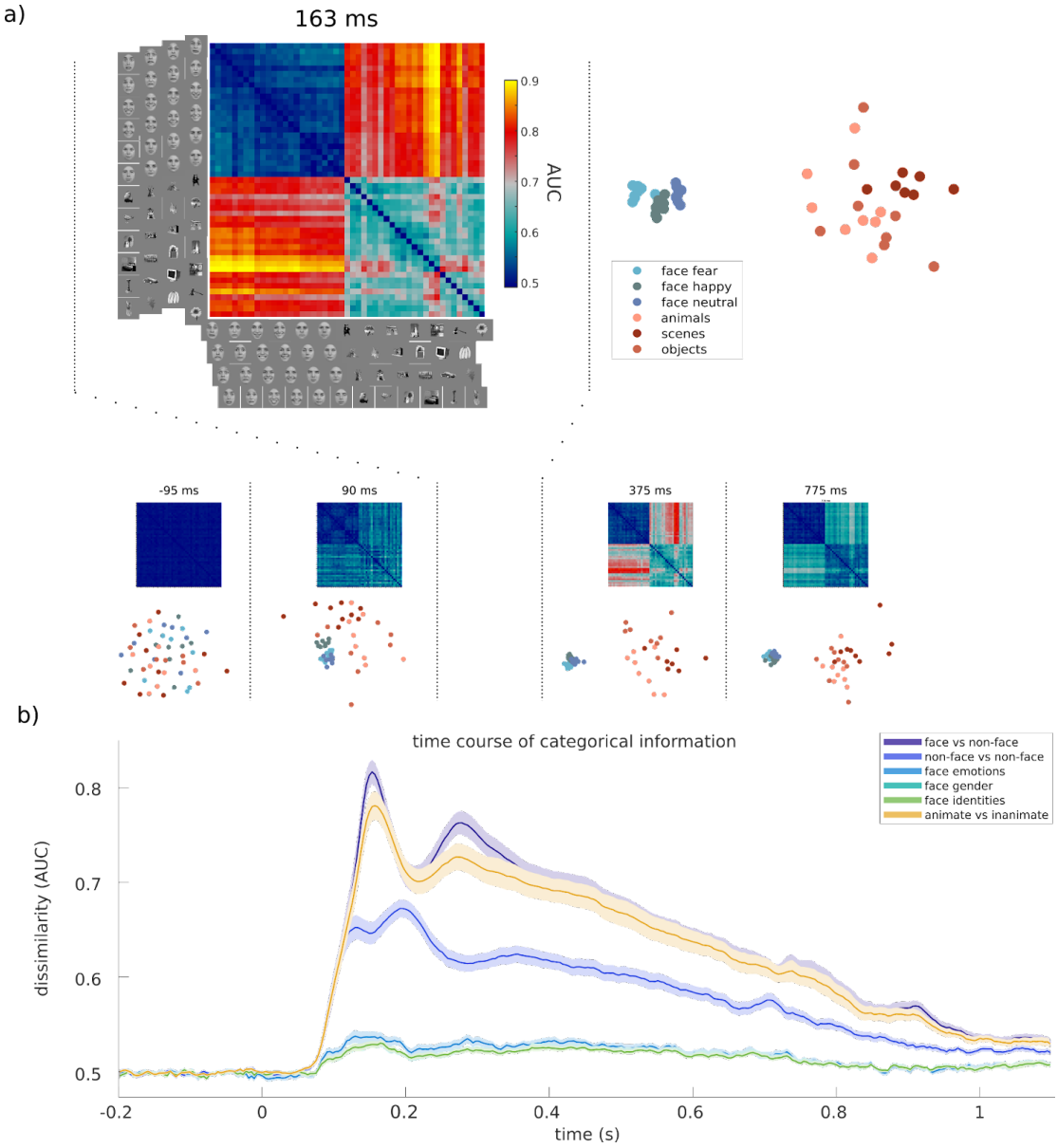


Figure 1.5. Dynamiques de la géométrie représentationelle en EEG. a) La Representational Similarity Analysis a été appliquée au signal EEG à travers le temps (provenant de l'article 4) en utilisant les patrons d'activité cérébraux multivariés comme mesure de dissimilarité entre les paires d'images pour créer des matrices de dissimilarité représentationnelle (*Representational Dissimilarity Matrices*, RDM). Une visualisation 2D de ces RDMs cérébrales à haute dimension montre des distinctions claires entre les différentes catégories (par exemple, les groupes de différents visages en bleu, les des scènes, animaux, etc. en teinte de rouges. b) Le décours de cette information catégorielle dans le temps (par exemple, les visages par rapport aux objets sans visage) est montré pour différentes catégories de stimulus. Les représentations cérébrales pour la distinction entre les visages et des objets (une caractéristique du N170, (Rossion &

Jacques, 2012), par exemple, domine toutes les autres distinctions catégorielles (Carlson et al., 2013 ; Kaneshiro et al., 2015) et atteint un maximum à 153 ms.

1.4 Réseaux de neurones artificiels et computations cérébrales

Si dans l'absolu les bénéfices qu'auront les modèles sur les neurosciences cognitives restent encore débattus (Schyns et al., 2022; Thompson, 2021), une chose est certaine: l'apprentissage automatique (*machine learning*) et l'apprentissage profond (*deep learning*) ont actuellement un impact épistémologique majeur, que ce soit par le développement des techniques d'analyses ou de nouveaux cadres explicatifs (Doerig, Sommers, et al., 2022; Kriegeskorte, 2015a; Richards et al., 2019) en cognition, en neurosciences, en neurosciences cognitives. Une nouvelle discipline, la neuroscience cognitive computationelle, de même qu'une nouvelle conférence éponyme active depuis quatre ans seulement (Kriegeskorte, 2015b; Naselaris et al., 2018) démontrent clairement l'état de l'expansion de l'utilisation de l'apprentissage profond dans ces disciplines.

Centraux aux neurosciences cognitives computationelles et à cette thèse sont les modèles d'apprentissage profond de reconnaissance d'objets. Ces modèles de reconnaissance d'objets sont majoritairement des modèles de'apprentissage profonds convolutifs (*convolutional neural networks*; CNNs), c'est-à-dire qu'ils utilisent un certain nombre de couches avec des filtres — des neurones artificiels — de différentes tailles et ayant des fonctions non-linéaires effectuant des convolutions sur leur input. Ces modèles incorporent typiquement un traitement hiérarchique à travers des couches effectuant différents types d'opérations de plus en plus abstraites qui sont, étonnamment, similaires à celles observées dans le système visuel ventral (Güçlü & van Gerven, 2015; Richards et al., 2019).

AlexNet (Krizhevsky et al., 2012) est un exemple de CNN *feedforward* bien connu avec une structure hiérarchique relativement simple (8 couches vs. modèles récents pouvant atteindre >100 couches; He et al., 2016). L'architecture de ce CNN est décrite à la figure 1.6. Même s'il a été développé en 2012 et est un peu moins performant que les derniers CNNs de reconnaissance d'objets, AlexNet a été montré plusieurs fois comme un des modèles se rapprochant le plus des

représentations du système cérébral ventrale (Richards et al. 2019; Khaligh-Razavi and Kriegeskorte 2014; Lindh et al. 2019; Allen et al. 2022; Yamins and DiCarlo 2016).

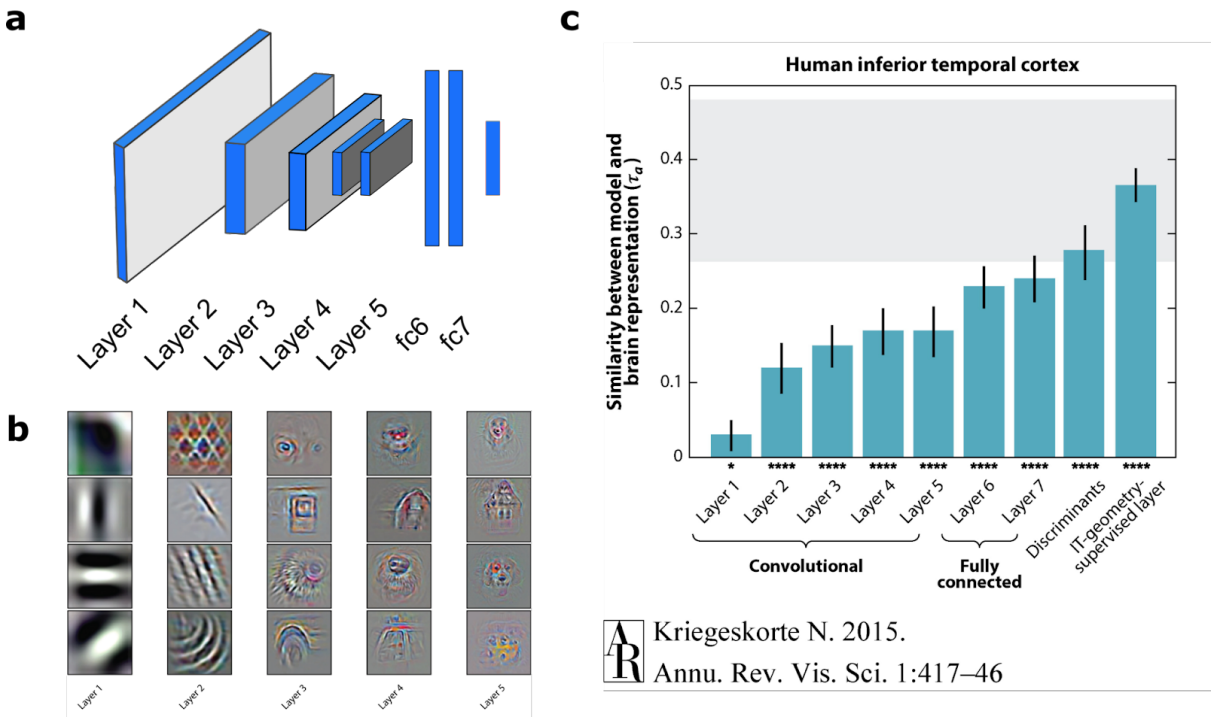


Figure 1.6. **Modèles d'apprentissages profonds de reconnaissance d'objets et associations avec représentations cérébrales. (a)** Le schéma montre le modèle d'apprentissage profond AlexNet (Krizhevsky et al., 2012), un *Convolutional Neural Network* composé de 5 couches convolutionnelles — conv1 (96 x 55 x 55), conv2 (256 x 27 x 27), conv3 (384 x 13 x 13) et conv5 (256 x 13 x 13) — ainsi que de trois couches *fully connected* — fc6 (4096 x 1), fc7 (4096 x 1) et fc8 (1000 x 1) — qui utilisent une fonction *softmax* pour transformer les caractéristiques en probabilité de classe (i.e. différents types d'objets). **(b)** Ce modèle d'apprentissage profond de reconnaissance d'objets montre des représentations visuelles de niveaux d'abstractions graduellement plus complexes dans les couches plus profonde de sont architecture, de simples grilles sinusoïdales à des parties d'objets et des objets entiers. Image provenant de (Güçlü & van Gerven, 2015) et adapté dans (Kriegeskorte, 2015b). **(c)** La similarité des représentations à travers les différentes couches d'AlexNet, obtenue par RSA selon des données de (Khaligh-Razavi & Kriegeskorte, 2014), montre une prédiction graduellement meilleure des représentations de la région cérébrale de haut-niveau hIT. Adaptée de Kriegeskorte, (2015b)**.**

Il fait ainsi partie des meilleurs modèles pour la similarité de ses computations avec celles du cerveau — i.e. il a un excellent *brainscore* (Nonaka et al. 2021). Une caractéristique importante de

ce type de modèles, et d'AlexNet spécifiquement, est qu'ils représentent des caractéristiques visuelles d'une complexité et d'une abstraction progressivement plus élevées le long de chacune de ses 8 couches (Güçlü & van Gerven, 2015). Regardons par exemple la figure 1.6b, qui illustre visuellement les représentations des différentes couches de ce CNN; depuis les caractéristiques de bas niveau (c.-à-d. le contraste, grilles sinusoïdales), jusqu'aux caractéristiques de niveau moyen (c.-à-d. Des *combinaisons* de grilles sinusoïdales, le contour, la forme, la texture) et aux caractéristiques de haut niveau (c.-à-d. parties d'objet et objets entiers). Ces modèles montrent ainsi un accès privilégié à des représentations ressemblant à celles du cerveau. Long et al. (2018), par exemple, ont montré que les représentations des caractéristiques de niveau intermédiaire, qui contribuent à une grande partie de l'organisation du système visuel ventral, culminent avec le traitement de la couche 4 de niveau intermédiaire d'AlexNet. Ces caractéristiques de niveau intermédiaire ont également été récemment associées à l'accès conscient aux informations visuelles (Lindh et al., 2019). Une autre raison d'utiliser ce modèle est sa relative simplicité par rapport à d'autres CNN plus récents (qui peuvent atteindre >100 couches; He et al., 2016). Plus généralement Khalig-Rasavi et al. (2014; voir figure 1.6c) ont montré, à l'aide de la RSA, que les couches de ce CNN expliquaient graduellement mieux l'activité cérébrale d'hIT, i.e. les couches profondes avec des représentations plus abstraites sont plus similaires à hIT (voir aussi Cadieu et al., 2014; Nonaka et al., 2021).

Ces modèles, tout comme le cerveau, effectuent une série de transformations qui passent graduellement de représentations basées sur la forme simple, à des représentations visuelles complexes, jusqu'au niveau de categorisation basique (e.g. un chat, une table). Des études récentes (Popham et al., 2021), dont celles décrites ici (article 4 et 5), ont utilisés des modèles d'apprentissage profonds *sémantiques* qui peuvent aller au-delà des représentations qui sont typiquement visuo-perceptives afin de sonder les représentations d'encore plus haut niveau. Ces modèles ont une architecture différente des CNNs, et, de façon plus importante, traitent des *phrases descriptives* — et non pas des images. Le Google Universal Sentence Encoder (Cer et al., 2018), par exemple, transforme une phrase (input) en un espace sémantique de 512 dimensions qui prédit plusieurs jugements humains de haut niveaux (Cer et al., 2018). L'arrivée de ces nouveaux modèles

computationnels est intéressante dans l'étude des différences individuelles puisqu'ils nous permettent de formaliser différents niveaux de traitement, i.e. visuel bas niveau, niveau intermédiaire, haut niveau— jusqu'au traitement sémantique des phrases. À l'heure actuelle, aucune étude n'a investigué le lien entre la similarité individuelle des représentations cérébrales à ces différents modèles et des variations perceptives.

1.5 Objectifs généraux et présentation des articles

Cette thèse comporte cinq articles principaux qui seront présentés dans cette section. L'auteur de cette thèse est le premier auteur de tous ces articles. Également inclus dans cette thèse se retrouvent un 6ième (annexe I) et un 7ième (annexe II) et 8ème article (annexe III), publiés pendant la même période et partiellement reliés à cette thèse. Chacun des articles principaux étudie les mécanismes perceptifs (articles 1, 2, 3) ou cérébraux (articles 4, 5) des différences individuelles en reconnaissance visuelle, tous dans l'optique de mieux répondre à une question simple : *pourquoi certains individus sont-ils meilleurs à reconnaître des visages?*

L'ensemble du travail sur l'implémentation cérébrale de la catégorisation, du code représentationnel et du contenu représentationnel mentionné ci-dessus a établi, depuis une vingtaine d'années, des mesures riches et sensibles rendant compte de la complexité du système visuel. D'un côté, les approches multivariées ont été utilisées pour investiguer *quand* et *comment* le cerveau effectue des processus importants sur son environnement visuel. La technique de *Representational Similarity Analysis* (RSA), par exemple, permet de comprendre comment certaines régions cérébrales encodent l'information visuelle, et permet également d'associer le traitement cérébral aux opérations effectuées par des modèles computationnels. Ces techniques sont toutefois agnostiques au contenu représentationnel. Les techniques psychophysiques de *reverse corrélation* permettent quant à elles de décrire le contenu de ces représentations (c.-à-d. l'information visuelle spécifique qui sous-tend la prise de décision perceptive). Cette thèse proposera ainsi cinq articles qui traiteront tour à tour du contenu représentationnel (psychophysique), des dynamiques cérébrales (EEG haute-densité), du code représentationnel (RSA), et de la similarité de ce code cérébral avec des

computations de différents niveaux d'abstractions (réseaux de neurones artificiels visuels et sémantiques). Afin d'associer ces mécanismes aux différences d'habiletés en reconnaissance faciale, ces techniques ont été utilisées pour sonder le cerveau d'individus provenant de populations extrêmement diversifiées: tant de population cliniques (i.e. individus sur le spectre de la schizophrénie [article 1], un cas important de prosopagnosie acquise [article 5]), que des individus neurotypiques [articles 2, 3, 4] et des individus avec des habiletés exceptionnels (i.e. des « *super-recognisers* », dans le top 1% du spectre d'habileté en reconaissance faciale [article 4]). Ces différentes techniques et populations contribueront ainsi à répondre à trois questions centrales aux neurosciences cognitives et computationnelles : **Quelle** information à propos des objets/visages est traitée, **comment** elle est transformée dans le cerveau et **quand** ces transformations se produisent.

1.5.1 Premier article

L'article 1 de cette thèse a été publié en 2020 dans la revue libre accès *Nature Schizophrenia* (maintenant appelée "*Schizophrenia*"; IF = 5.044). Cet article décrit la première caractérisation du contenu représentationnel d'individus avec un diagnostic comorbide de spectre de la schizophrénie et d'anxiété sociale (SZ&SAD). Ceci a été accompli à l'aide d'une technique psychophysique de *reverse correlation* déjà bien établie, *Bubbles (Gosselin & Schyns, 2001)*. Même si les déficits *sociaux* des individus SZ&SAD sont particulièrement sévères (e.g. isolement social, augmentation du taux de suicide, moins bonne qualité de vie générale), les mécanismes de reconnaissance des expressions faciales des individus avec ce trouble comorbide ont été pratiquement inexplorées. En collaboration avec des chercheur(e)s clinicien(ne)s, nous avons soumis cette population à un nombre important d'essais d'une tâche Bubbles pendant qu'ils reconnaissaient différentes expressions faciales (peur, joie, colère, et neutralité), ce qui nous a permis de faire trois découvertes importantes. Nous avons: i) révélé pour la première fois des déficits d'habiletés en reconnaissance faciale des expressions chez cette population; ii) associé directement ces déficits à des symptômes fonctionnels (symptômes négatifs & d'anxiété sociale); et iii) nous avons pu montrer que leurs déficits d'habiletés perceptives provenait de leur utilisation anormale et inefficace de l'information visuelle d'un visage. En effet, les individus avec SZ&SAD n'utilisent pas la région des yeux pour categoriser

les expressions faciales, sous-utilisent les détails fins (hautes fréquences spatiales), et sur-utilisent les détails grossiers (basses fréquences spatiales). Cette étude décrit ainsi pour la première fois les mécanismes perceptifs responsables de la confusion fréquente des visages exprimant la colère et la peur chez cette population. Ces résultats ont également des implications cliniques importantes — un programme d'entraînement en reconnaissance faciale visant à réduire ces confusions chez les individus sur le spectre de la schizophrénie, par exemple, est en cours d'élaboration.

1.5.2 Deuxième article

L'article 2 de cette thèse a été publié en 2021 dans la revue *Frontiers in Psychology* (IF = 4.23). Cet article décrit l'utilisation d'une nouvelle technique de *reverse correlation* permettant de révéler, simultanément, l'utilisation de trois informations clefs du système visuel humain (Alink & Charest, 2020) : la position (e.g. l'oeil gauche), les fréquences spatiales (la granularité de l'information visuelle de grossière à fine) et l'orientation (e.g. les informations horizontales). Nous avons appliqué ce nouvel outil psychophysique afin de cartographier les caractéristiques faciales nécessaires à une tâche de reconnaissance importante et très peu étudiée, la classification de l'âge des visages. Ce projet réalisé en collaboration avec des chercheures en psychologie médico-légale (J. Woodhams et J. Kloess; *University of Birmingham*) visait, également, à établir les bases de la classification automatique des images faciales dans les bureaux de police du Royaume-Uni. Nous avons révélé que les informations les plus importantes pour catégoriser l'âge des visages sont les informations horizontales de granularité moyenne dans les régions des yeux et les informations verticales dans la région correspondant à l'os nasale.

1.5.3 Troisième article

L'article 3 de cette thèse est en préparation (une version préliminaire du contenu de l'article a été présentée à la conférence annuelle de la Vision Science Society en 2019 et un résumé de cette présentation par affiche a été publié dans le Journal of Vision, (Faghel-Soubeyrand, Alink, Bamps, Gervais, et al., 2019). Il sera soumis à la revue *Psychological Science* (IF = 7.29). Cet article

modélise la capacité du système visuel de 120 participants à modifier leurs représentations faciales internes en fonction de différentes tâches de reconnaissance faciale, et ce afin de mieux comprendre la relation entre cette capacité et l'habileté perceptive individuelle.

Spécifiquement, ce travail a abordé deux questions importantes pour une compréhension écologique de l'habileté en reconnaissance faciale, mais qui ont été négligées jusqu'à présent. En effet, les études publiés à ce jour sur les différences individuelles en reconnaissance faicale ont étonnamment omit le fait que les objectifs de notre système visuel doivent constamment changer pour extraire différentes informations sur les visages, par exemple de l'information sur l'identité, le sexe ou les expressions faciales. La première question concerne ainsi la généralisation de la capacité d'une personne à traiter différentes informations faciales : les individus avec une bonne habileté sont-ils performants à travers diverses tâches de reconnaissances ? Nous avons trouvé des preuves en ce sens. Nous avons observé, dans un échantillon de 120 participants, de fortes corrélations entre l'habileté individuelle à reconnaître le sexe & l'habileté à reconnaître les expressions faciales, ainsi que de fortes corrélations entre leur capacité à reconnaître le sexe et leur capacité à identifier les visages. La deuxième question concerne les mécanismes spécifiques qui sous-tendent cette habileté à travers les différentes tâches : quelles stratégies perceptives les individus performants utilisent-ils pour parvenir à une bonne performance entre les tâches ? Nous avons abordé cette question en mesurant, à l'aide de technique de *reverse correlation* utilisée dans l'article 2, les changements dans les représentations perceptives des visages à travers deux tâches — (1) reconnaissance des expressions et (2) du sexe des visages — chez la même cohorte de 120 individus. Nous avons enregistré plus d'un demi-million d'essais (2400 essais par sujet et par tâche, pour un total de 576 000 essais) avec cette technique, ce qui nous a permis d'associer le contenu visuel avec l'habileté en reconnaissance faciale chez ces mêmes individus. Cet ensemble massif de données nous a permit de découvrir que les individus avec une meilleure habileté perceptive s'appuient sur un contenu perceptuel *plus similaire d'une tâche à l'autre*. Ces résultats fournissent les premiers mécanismes expliquant l'habileté d'un individu *à travers les tâches* de reconnaissance faciale, et suggère l'utilisation d'une stratégie générale/unique à travers les différentes demandes. Nous proposons qu'un tel mécanisme s'explique par un compromis entre trois impératifs internes et externes : i) notre

appareil cognitif/cérébral limité, ii) nos objectifs visuels diversifés et en constante évolution, et iii) les propriétés physiques requises pour extraire ces différentes informations des visages.

1.5.4 Quatrième article

L'article 4 de cette thèse est *under review* à *Nature Communications* (IF = 17.69). Dans ce projet collaboratif entre cinq laboratoires internationaux (Suisse, Royaume-Uni et Canada), nous avons caractérisé, pour la première fois, la nature des représentations cérébrales covariant avec la capacité à reconnaître des visages en utilisant des modèles computationnels visuels et sémantiques. Une contribution majeure de cet article a été la démonstration que l'habileté en reconnaissance faciale affecte des processus qui vont jusqu'aux représentations sémantiques, c'est-à-dire au-delà des marqueurs perceptifs (visuels) traditionnellement associés à la reconnaissance de visages.

Il est, en effet, envisageable que des différences de contenu représentationnel, mais aussi de la géométrie représentationnelle, sous-tendent les différences individuelles au niveau perceptuel. Il est possible, par exemple, que deux individus représentent les mêmes détails visuels pour discriminer une girafe d'un corbeau, mais que l'un encode plus efficacement ces différences (e.g. à travers des connexions neuronales plus rapides). Il est aussi plausible qu'un *super-recognizer* (i.e. un individu avec une habileté extraordinaire à mémoriser/reconnaître les visages; (Russell et al., 2009) possède un code représentationnel permettant de distinguer davantage les différents visages qu'un individu typique. De telles représentations plus séparables (Grill-Spector & Weiner, 2014) n'ont, par contre, jamais été observées chez des individus différant au niveau de leur habileté perceptive. Qui plus est, nous ignorons si des individus *super-recognizers* ont un système visuel optimal spécifiquement pour les visages (c.-à-d. spécifique aux catégories), ou si cet avantage s'étend à d'autres objets/scènes (c.-à-d. général).

L'article 4 de cette thèse a évalué les différences cérébrales fonctionnelles et le code représentationnel (Kriegeskorte et al., 2008) de cerveaux des *super-recognizers*, et des individus typiques. Nous avons enregistré une grande quantité de données d'électroencéphalographie à haute-densité (>100 000 essais au total, ~120 heures d'enregistrements de signaux cérébraux) en

réponse à des stimuli provenant de plusieurs catégories de base (visages, scènes, objets et animaux) chez 16 super-recognisers (i.e. des individus se situant dans le 1% supérieur des capacités de reconnaissance des visages) et 17 individus typiques (i.e. variant typiquement sur leur habileté perceptive). Nous avons utilisé la RSA et des outils d'apprentissage machine de pointe pour étudier cet ensemble de données cérébrales et neuropsychologiques massives. Nous arrivons à trois conclusions importantes : (i) l'habileté individuelle en reconnaissance faciale peut être décodée avec une haute précision (~80 % de précision) — et ce à partir d'une seule seconde d'enregistrement de l'activité cérébrale; (ii) cette prédiction est possible lorsque les participants traitent des visages mais aussi des objets visuels qui ne sont pas des visages, ce qui confirme que des mécanismes généraux dans le cerveau humain sont en partie responsables des différences individuelles en reconnaissance faciale; et (iii) cette prédiction de l'habileté en reconnaissance faciale à partir de l'activité cérébrale est possible en raison de différences au sein des représentations visuelles de niveau d'abstraction intermédiaires (tel que caractérisées par des modèles d'apprentissage profond entraînés à la reconnaissance des objets) et des représentations sémantiques (tel que dérivées par des modèles d'apprentissage profond du langage). Cette association entre les processus sémantiques et la capacité à reconnaître les visages avait été postulée dans certains modèles en reconnaissance faciale, mais elle n'avait jamais été observée empiriquement jusqu'à présent.

Un article de vulgarisation portant sur ce travail de recherche et écrit par l'auteur de cette thèse est disponible en ligne au lien suivant: Décoder le cerveau de super-physionomistes | IVADO.

1.5.5 Cinquième article

L'article 5 de cette thèse est *under review* à Cerebral Cortex (IF = 5.4). Ce projet décrit la première investigation neuro-computationelle de la patiente PS, considéré comme le cas le plus spécifique et le plus étudié de prosopagnosie acquise (Rossion, 2022b). La patiente PS présente des troubles importants en reconnaissance faciale d'identité; ces déficits se sont manifestés suite à des lésions cérébrales relativement focalisées dans le système de reconnaissance faciale (partie postérieure du cortex occipitotemporal [OFA] droit, partie ventrale du cortex occipitotemporal [FFA] gauche, et petite partie du gyrus temporal médian droit). Nous avons enregistré les signaux cérébraux de PS et

de 19 contrôles neurotypiques, dont quatre appariés sur plusieurs facteurs neuropsychologiques et démographiques, à l'aide de l'EEG haute-densité et des méthodes semblables à celles de l'article 4 et mentionnées ci-dessus. Nous avons d'abord utilisé l'apprentissage machine (modèles de classifications de type *linear discriminant*) pour extraire le décours temporel de l'information cérébrale i) d'identité faciale et ii) d'identités de "non-visages" (i.e. d'autres catégories visuelles telles que des scènes, animaux, etc.) chez PS et les contrôles. Ces outils ont été capables de révéler une dissociation entre l'évidence cérébrale d'identité faciale et d'identité des objets chez PS qui calquent ses troubles perceptifs spécifiques aux visages. Ensuite, nous avons utilisés les mêmes analyses multivariées (i.e. RSA et *decoding*) et les mêmes modèles computationnels décrits dans l'article 4 (réseaux de neurones artificiels profonds visuels et langagiers) pour modéliser les signaux cérébraux de PS. Cette première modélisation computationnelle du cerveau de PS a permis de faire plusieurs découvertes importantes. D'abord, nous révélons que la prosopagnosie est associée à une réduction de la qualité des représentations visuelles de relativement haut-niveau, ainsi qu'à une réduction encore plus significative des opérations cérébrales sémantiques de plus haut niveau d'abstraction (i.e. provenant de descriptions écrites d'images). Cette réduction de la qualité des représentations sémantiques est présente tôt dans le cerveau, atteint un sommet autour de la composante N170, et persiste ensuite jusqu'à des composantes cérébrales plus tardives. Une analyse plus approfondie de la dynamique cérébrale de PS a révélé que le code neuronal de la patiente semble plus stable dans le temps par rapport aux témoins, avec moins de changements entre les représentations cérébrales précoces et tardives.

Chapitre 2 – Article 1

Abnormal visual representations associated with confusion of perceived facial expression in schizophrenia with social anxiety disorder

Faghel-Soubeyrand[1,2], S., Lecomte[1] T., Bravo[1], M.A, Lepage[3], M., Potvin[4], S., Abdel-Baki[5], A., Villeneuve[6], M., & Gosselin[1], F.

1. Département de Psychologie, Université de Montréal, Canada
2. School of Psychology, University of Birmingham, UK
3. Department of Psychiatry, McGill University, Canada
4. Département de Psychiatrie, Université de Montréal, Canada
5. Centre hospitalier de l'Université de Montréal-Hôpital Notre-Dame, Canada
6. Institut universitaire en santé mentale de Montréal, Canada

Corresponding author: Simon Faghel-Soubeyrand,
simon.faghel-soubeyrand@umontreal.ca

Publié : Faghel-Soubeyrand, S., Lecomte, T., Bravo, M. A., Lepage, M., Potvin, S., Abdel-Baki, A., Villeneuve, M., & Gosselin, F. (2020). Abnormal visual representations associated with confusion of perceived facial expression in schizophrenia with social anxiety disorder. *NPJ Schizophrenia*, *6*(1), 28. https://doi.org/10.1038/s41537-020-00116-1

Word counts. Abstract: 276. Total word count : 4381 (main text : 3073, method section : 1308) .

Abstract

Deficits in social functioning are especially severe amongst schizophrenia individuals with the prevalent comorbidity of social anxiety (SZ&SAD). Yet, the mechanisms underlying the recognition of facial expression of emotions—a hallmark of social cognition—are practically unexplored in SZ&SAD. Here, we aim to reveal for the first time the visual representations SZ&SAD individuals (n=16) and controls (n=14) rely on for facial-expression recognition. To do so, we ran a total of 30,000 trials of a facial-expression categorisation task (fear, happy, anger and neutral) with Bubbles, an established data-driven technique. Results showed that SZ&SAD's ability to categorise facial-expression was impaired compared to controls. More severe negative symptoms (flat affect, apathy, reduced social drive) was associated with more impaired emotion recognition ability, and with more biases in attributing neutral affect to faces. Higher social anxiety symptoms, on the other hand, was found to enhance the reaction speed to neutral and angry faces. Most importantly, Bubbles results showed that these abnormalities could be explained by inefficient visual representations of facial expression of emotions. Indeed, compared to typical controls, SZ&SAD subjects relied less on fine facial cues (high spatial frequencies) and more on coarse facial cues (low spatial frequencies). Moreover, SZ&SAD participants never relied on the eye regions (only on the mouth area) to categorize facial expression of emotions. We argue that schizophrenia *and* social anxiety disrupt visual emotion representations to a much greater extent than either schizophrenia or social anxiety diagnostics on their own.

Key words: schizophrenia, social anxiety disorder, emotion face perception, visual representations, spatial frequency, emotion misattribution

Introduction

Deficits in social cognition, notably in the recognition of facial expression of emotions, are one of the hallmark impairments in individuals with schizophrenia (SZ) [1]. Social cognition deficits are significantly related to a person's level of functioning [2], and emotion perception in particular is strongly and consistently associated with various domains of functioning, such as community functioning, social behaviors, social skills, and social problem solving [3].

Schizophrenia is also associated with multiple comorbid disorders, including anxiety disorders and more specifically social anxiety disorder (SAD) [4,5]. Several studies have demonstrated that not only does SAD worsen the prognosis of schizophrenia, but also impedes the overall social recovery of the individual [6]. Furthermore, individuals suffering from comorbid schizophrenia and social anxiety (SZ&SAD) have a higher rate of suicide attempts, a high prevalence of alcohol/substance abuse disorders, and worse quality of life compared to those with schizophrenia without social anxiety [7]. Thus, considering the serious consequences of comorbid SZ&SAD, better understanding the mechanisms underlying their emotion recognition deficits is important.

To our knowledge, only two studies have examined emotion recognition and confusion patterns in SZ&SAD [8,9], and they have led to contradictory results. The current study's main objective is to resolve this conflict and to shed light on the perceptual mechanisms responsible for the categorization of facial expression of emotions in SZ&SAD. To do so, we used Bubbles[10], an established data-driven method, which enabled us to reveal the specific visual representations (i.e. the precise facial features and spatial-frequency information) that individuals with SZ&SAD rely on to decode facial expression of emotions [11–13]. This also allowed us to measure SZ&SAD's emotion recognition ability with robust psychophysical metrics, to describe their emotion

confusion patterns, and to reveal any association between these facial-affect processing metrics and their psychiatric symptoms. Given the absence of studies on the visual representations in individuals with SZ&SAD and even in SAD, the following paragraphs will focus on reviewing findings relevant to the categorization of facial expression of emotions in individuals with SZ.

In individuals on the schizophrenia spectrum, the presence of facial expression of emotion categorization deficits is now well-established. In an attempt to explain these impairments, clinical and cognitive psychologists have recently proposed that the emotion perception abnormalities of SZ may emerge from over-attributing specific emotions to affective faces, specifically negative emotions such as anger, fear or disgust [14,15] [16–18]. Likewise, psychologists have speculated that socially anxious individuals should also show a bias for negative emotions because of their fear of being judged in a negative way by their peers [19,20]. Empirical studies in both clinical and non-clinical populations have revealed some associations between a bias for negative emotions and social anxiety traits [11,21–26]. As far as we know, the joint effect of schizophrenia *and* social anxiety on categorization biases has only been explored in one study [9]. The authors observed that the presence of social anxiety in SZ resulted in patients having more difficulty recognizing neutral faces, confusing them with faces expressing other emotions, compared to those with schizophrenia and no social anxiety.

Prior work with individuals with SZ using behavioral and brain imaging methods has revealed low-level perceptual abnormalities that may be linked to these facial expression categorization impairments and biases. Indeed, individuals with SZ tend to have reduced contrast sensitivity to low-spatial frequencies gratings [27–29] (LSF; coarse visual information) and reduced sensitivity to LSF filtered faces [30,31,32]. These studies as well as others [33,34] have promoted the idea of a magnocellular deficit in SZ [35]. Using similar methods, researchers have also proposed that deficiencies at a higher-level of the visual system could explain poor facial-affect processing in SZ. It is thought, for example, that top-down modulation processes such as the “coarse-to-fine”

object/face recognition process [36,37] could be linked to the emotion processing impairments of SZ patients [38,39]. Arguing for deficits on such top-down processes, recent brain imaging studies[39] have proposed that individuals with SZ may compensate for their poor processing of LSF (coarse) information with a visual strategy relying more on HSF (fine) information. Quite surprisingly, however, even though these past studies have speculated on the high-level internal representations SZ rely on in various face-recognition tasks, only two studies have looked directly at such visual representations in SZ [12,13]. These two studies found atypical visual representations in SZ (e.g. less reliance on the eye region), but no evidence for a bias toward HSF in SZ. Again, no one has ever explored the effect of clinically diagnosed social anxiety[11], nor the joint effect of schizophrenia and clinical social anxiety on visual representations.

The objectives of the present study were thus the following: 1) to measure the general ability for facial-expression recognition in individuals with SZ&SAD; 2) to examine categorization biases in this population and to explore their link with psychiatric symptoms; and 3) to reveal the visual memory representations underlying the facial expression of four basic emotions in SZ&SAD.

Method

Participants and procedure

A total of 30 participants were recruited for this study. The first group consisted of 16 individuals diagnosed with both a schizophrenia spectrum (schizophrenia, n = 9; schizo-affective disorder, n = 7) and comorbid social anxiety disorder. The second group was composed of 14 normal controls (see Table 1 for socio-demographic and clinical information). Clinical participants were referred and diagnosed by the attending psychiatrist (according to the DSM-5 criteria), by means of the first episode psychosis clinic of the *Institut Universitaire en Santé Mentale de Montréal* (IUSMM), the *Centre Hospitalier de l'Université de Montréal* (*CHUM)* and the *Douglas institute*

Centre for Personalized Psychological Intervention for Psychosis. The normal control group was composed of neurotypical individuals, i.e. individuals who did not have any psychological or neurological disorder. All participants provided written consent to take part in the study. This study was approved by the Ethics and Research Committee of the IUSMM, CHUM, Douglas Institute and the ethics committee of the Université de Montréal.

All participants were administered a socio-demographic survey, and performed the *Bubbles Facial Emotion Recognition Task* (henceforth called *Bubbles experiment*). In addition, clinical participants were evaluated with the following instruments: 1) *Structured Clinical Interview for DSM-5* (SCID[40]), which allowed to confirm patient diagnosis according to the DSM-5 criteria; 2) *Brief Psychiatric Rating Scale–Expanded Version* (BPRS[41,42], which consists of a 24-item semi-structured interview that assesses the severity of psychiatric symptoms and which was delivered by trained-to-gold-standard graduate students. The following dimensions were used for analyses: negative symptoms, positive symptoms, affect, and activity; 3) *Brief Social Phobia Scale* (BSPS[43]), which evaluates the severity of social anxiety and includes 18 self-report items answered on a five-point Likert scale (from 0 to 4). The scale has adequate psychometric properties[43]. The following subscales were generated: fear, avoidance and physiological activation.

Table 2.1. Sociodemographic and clinical information of participants

	SZ&SAD group (n = 16)	**Control group (n = 14)**	**Group differences**
Age	27.9 (4.4)	27.2 (4.4)	p = .590
Sex	43.8% women	33.3% women	p = .722
Years of education	12.9 (2.4)	16.8 (3.3)	p = .001
Age first consultation	20.9 (5.2)	—	—
No. hospitalizations	2.1 (2.1)	—	—
Brief Psychiatric Rating Scale			
Negative symptoms	9.6 (3.0)	—	—
Positive symptoms	9.6 (6.9)	—	—
Affect	12.8 (5.9)	—	—
Activity	9.8 (2.7)	—	—
Total score	41.8 (12.8)	—	—
Brief Social Phobia Scale			
Fear	16.6 (6.3)	—	—
Avoidance	16.5 (5.6)	—	—
Physiological	5.9 (3.3)	—	—
Total score	39.1 (12.2)	—	—

Bubbles experiment

Stimuli. 40 grayscale face stimuli (256 x 256 pixels; approximately 5.72 x 5.72 deg of visual angle) from the STOIC database ([5 females & 5 males] * 4 facial expressions[12,13,44,45]) were used as base faces. The emotions (happy, fearful, angry & neutral) conveyed by these stimuli are expressed by trained actors and were selected to be highly recognizable.

Experimental procedure. To map which parts of expressive faces at different spatial scales were used by individuals with SZ&SAD and controls, we used the Bubbles method and an emotion categorization task (joy, fear, anger neutrality). This sampling procedure is illustrated in

Figure 2.1 (for further details, see [10]). Briefly, faces were drawn randomly from the 40 exemplars as well as their mirror-flipped images, and randomly sampled using Gaussian apertures (bubbles) in the 2D image plane and in five spatial frequency bands. Spatial filtering was achieved with Laplacian pyramid transforms [46]. The quantity of visual information (i.e. the number of bubbles) was adjusted throughout the task by the QUEST algorithm [47] so that each subject maintained an accuracy of 75%. This quantity of visual information was our measure of face emotion categorization performance (see data analysis section). These stimuli were shown at the center of the screen and remained there until the subject pressed a key corresponding to one of the four possible emotions. Every participant completed 10 blocks of 100 trials each.

Data Availability

The data that support the findings of this study are available on request from the corresponding author [S.F.S].

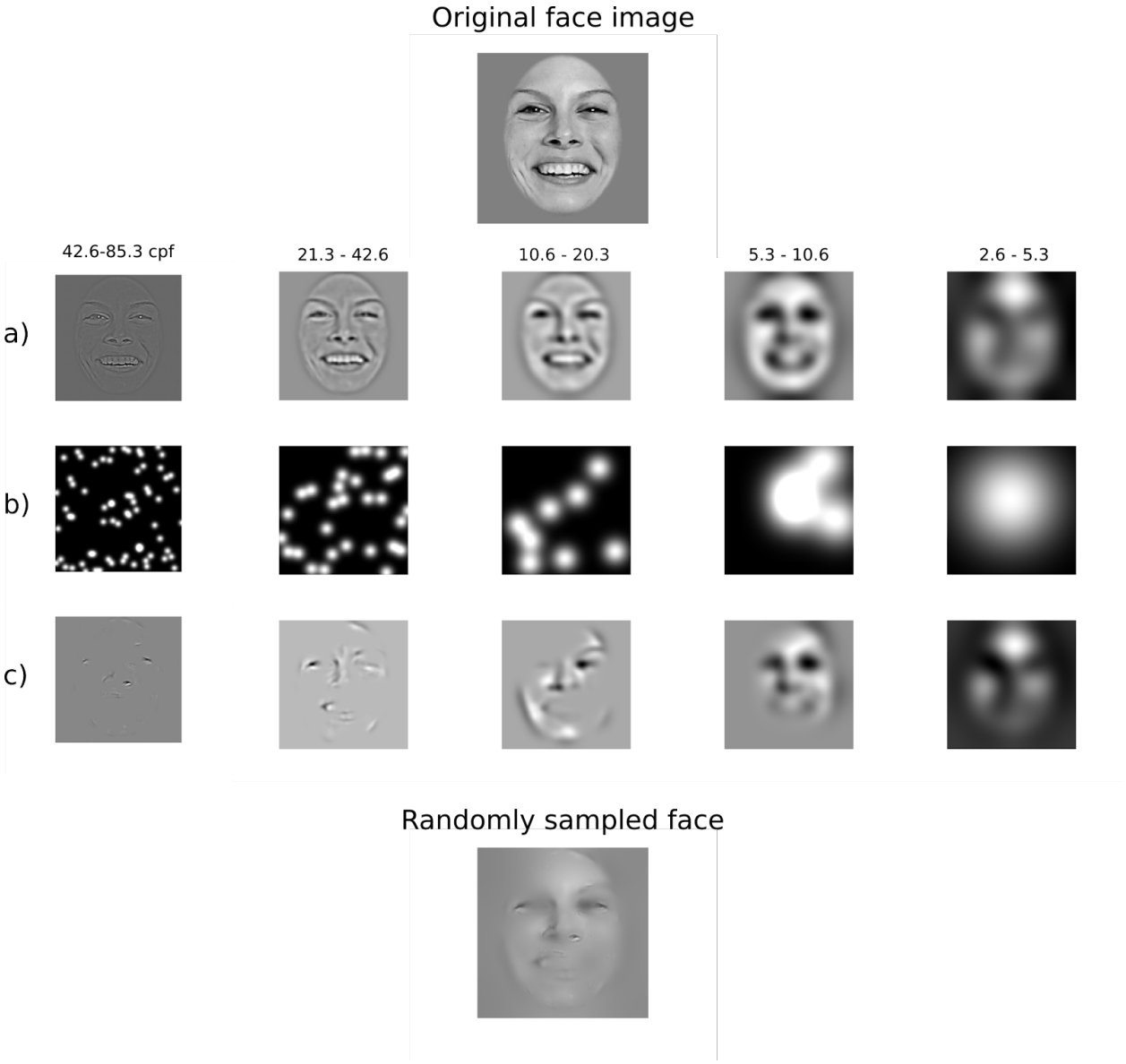


Figure 2.1. Creation of the "Bubblized" facial-expression stimuli. An original face image was decomposed into five spatial frequency bands via Laplacian transform (a). Small Gaussian apertures (i.e. the bubbles) were then placed at random locations, for each spatial frequency band separately (b). The information revealed by the bubbles for each SF band (c) was then fused across the five frequency bands to produce a single trial's experimental stimulus (bottom image).

DATA ANALYSIS

Association between emotion categorization and clinical psychopathology

To associate emotion recognition processes to clinical psychopathology, individual psychophysical metrics within the SZ&SAD cohort (e.g. individual facial-expression recognition performance thresholds) were Spearman-correlated with individual clinical scores (e.g. social anxiety trait score, as measured by the BSPS). Individual psychophysical performance thresholds for the emotion categorisation task [48,49] were computed from the best-fitted Gaussian cumulative distribution function (i.e. accuracy as a function of number of *bubbles* apertures across trials). Individual average reaction times for correct responses were computed for the four emotion categories, discarding outlier trials (trials with $Z > 2.5$, $Z < -2.5$ or < 200ms were rejected).

Confusions: association with clinical psychopathology

To reveal if SZ&SAD's clinical psychopathology were associated with specific confusion patterns between emotions, a 4x4 (categorized emotion x presented emotion) confusion matrix was first derived for each individual and averaged across participants. Specifically, we computed the probability that each observer classified a given signal (e.g. an angry face) as a specific emotion (e.g. neutral), for each possible categorization and signal type. To know whether and which specific confusion patterns were associated with clinical psychopathology, we Spearman-correlated the between-individual variation in psychiatric score of interest (BRPS & BSPS scores) with every categorization x facial-affect probability pair (e.g. an angry face confused as a neutral one). Multiple comparisons were controlled for using Bonferroni corrections.

Classification Images

Our main goal was to reveal the visual memory representations of individuals with SZ&SADs

and of controls when categorizing the facial expression of four basic emotions. Considering its use to explore complex memory representations, reverse correlation techniques prescribes a high number of trial repetition per participant. In this study, thus, robust within-participants measures (1000 trials per individual, for a total of >30 000 trials) was preferred over a large sample size [50,51]. To reveal these visual representations, we performed multiple linear regressions on the location of the bubbles and z-scored response accuracies across trials, for each subject, spatial frequency band and emotion. Each of these regressions produces a plane of regression coefficients that we call a classification image (CI); it reveals parts or features of faces—in their specific spatial frequency band—that are systematically associated with a participant's accurate emotion categorization. CIs were then summed across participants of each group per emotion and frequency band, smoothed with a Gaussian kernel of 15 pixels of standard deviation, and transformed into Z scores. Cluster tests [52] (search region=65,536 pixels; arbitrary z-score threshold = 2.7; cluster size statistical threshold=277 pixels; $p<.05$, two-tailed) were used to assess the statistical significance of the resulting group CIs. The Cluster test corrects for family-wise error rate while taking into account the spatial correlation that is inherent to the *Bubbles* method. To specifically test whether SZ&SAD relied more on high SF, we contrasted (unpaired t-test), between SZ&SAD and controls, the individual average z-scores contained in the face area for each of the five frequency bands.

Results

Emotion categorization performance

One SZ&SAD participant was excluded from further analysis because of his poor performance (number of bubbles threshold > $\underline{X} \pm 3\sigma$; $N_{SZ\&SAD}=15$). No control participants were excluded based on this criteria. The resulting total sample's (i.e. SZ&SAD and controls) average number of bubbles threshold ranged from a low of 46 (i.e. best performance) to a high of 231 (poorest

performance).

SZ&SAD performed worse—they required more bubbles—than neurotypicals to categorize facial expression of emotions ($M_{SZ\&SAD}$= 130.3471, M_{CTL}= 90.3380; $STDEV_{SZ\&SAD}$= 50.8760, $STDEV_{CTL}$= 34.3402 F(26,1)=5.75, p=.0239) and were slower than controls ($F_{group\_effect}(104,1)=18.8$, p<.0001; $F_{emotion\_effect}(104,3) = 4.37$, p=.006; $F_{interaction}(104,3) =0.22$, p>.3).

Association between emotion categorization accuracy and clinical psychopathology

The more negative symptoms (blunted affect, apathy, and reduced social drive) as assessed by the BPRS a person with SZ&SADs displayed, the worse this person was at categorizing facial expression of emotions ($r(14) = .5768$, p = .0244). There was no association between reaction time and BPRS symptoms (all *ps* >.05). In contrast, social anxiety symptoms as assessed by the total score of the BSPS were not significantly linked significantly to categorization performance ($r(14) = -.42$, p = .14), but they were linked to faster reaction times for both anger and neutral faces in participants with SZ&SAD ($r_{anger}(14) = -.63$, p=.01; $r_{neutral} = -.52$, p = .04). More specific (*post-hocs*) associations revealed that the fear dimension of social anxiety symptoms were driving faster categorization of anger faces ($r_{anger}(14) = -.53$, p=.0287; $r_{neutral}(14) = -.48$, $p=.07$), while the avoidance dimension was driving faster categorization of neutral faces ($r_{anger}(14) = -.43$, p=.10; $r_{neutral}(14) = -.64$, $p=.0109$).

Association between emotion confusions and clinical psychopathology

The SZ&SAD group confusion matrix (shown in Figure 2.2a, second row) revealed three significant frequent confusions during the categorization of facial expression of emotions: i) neutral faces were often miscategorized as “fearful”, ii) angry faces were often miscategorized as

"neutral", and iii) fearful faces were often miscategorized as "angry" ($Probability_{Observed}$ = 13.14%, χ^2=112.0896; $Probability_{Observed}$ = 13.14%, χ^2=112.0896; $Probability_{Observed}$ = 10.64%, , χ^2= 24.9799, $Probability_{Expected}$ = 8.3%, i.e. $\frac{(1-accuracyExpected)}{(nClass-1)}$, all *ps*<.05, Bonferroni-corrected). There were no significant contrasts between confusions of SZ&SAD and controls (see supp. material for further details and interpretation).

The correlation between individual biases in the decoding of specific facial expressions and the four BPRS dimensions described earlier is shown in Figure 2.2b. Congruent with our previous result, SZ&SAD individuals with more severe negative symptoms were more prone to miscategorize angry and fear faces as "neutral" ($r(14)_{angerAsneutral}$ = .7823; $r(14)_{fearAsneutral}$ = .7521, p<.003, Bonferroni-corrected). We also found a negative correlation between negative symptoms and accurate anger categorization ($r(14)$ = -.7594, p<.003, Bonferroni-corrected). Moreover, the higher an individual with SZ&SAD scored on the affect dimension (presence of depression, anxiety and guilt), the more this individual confused angry faces as joyful ($r(14)$ = .7756, p<.003, Bonferroni-corrected).

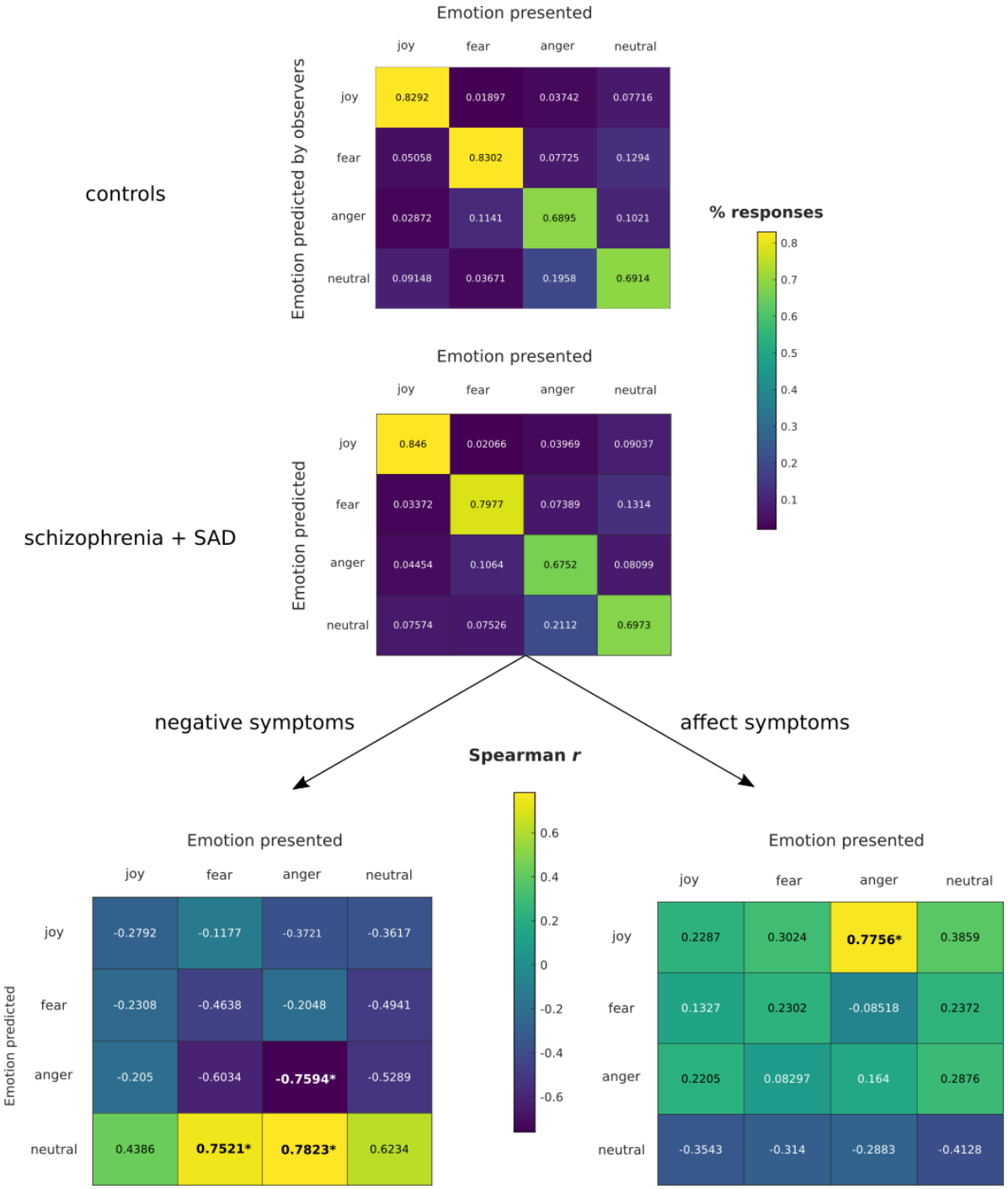


Figure 2.2. Confusion of facial expression of emotions in (a) controls and (b) SZ&SAD. (c) Correlation between the SZ&SAD confusion matrix and clinical negative symptoms. (d) Correlation between the SZ&SAD confusion matrix and clinical affective symptoms.

Visual representations for emotion categorisation

The classification images of control participants show that they used the mouth and, to a lesser extent, the eyes (mostly the left eye) to categorize facial expression of emotions (Figure 3a; see also [44]). SZ&SAD participants, however, only relied on the mouth region to categorize facial expressions.

We also examined the use of facial features at different spatial frequencies across emotions (see Figure 2.3b). Individuals with SZ&SAD (green) represented emotion cues from the mouth at spatial frequencies between 10.6 and 42.6 cpf. For controls (red), their visual representation of eye emotion cues contained only high spatial frequencies (>21.3 cycles per face [cpf]), while the mouth region contained only lower spatial frequencies (20.3-5.3 cpf).

We also examined the use of different spatial frequencies across emotions and facial features (see Figure 3b). This revealed that SZ&SAD relied less overall on face-emotion information of higher frequencies than controls (42.6-85.3 cycles per face [cpf]; $M_{SZ\&SAD}$= 0.0741, $M_{CONTROL}$= 0.1414, t (27) = -3.228, p = .0033), and used more low spatial frequency face-emotion information (2.6-5.3cpf; $M_{SZ\&SAD}$ = 0.2369, $M_{CONTROL}$= 0.0722, t (27) = 2.4518, p = .0210; mid-range SFs showed no significant contrasts). We finally examined the use of facial features at different spatial frequencies for all emotions (see Figure 2.3c-f). Individuals with SZ&SAD mostly represented emotion cues from the mouth at spatial frequencies between 10.6 and 42.6 cpf. For controls, their visual representation of eye emotion cues contained only high spatial frequencies (>21.3 cpf), while the mouth region contained only lower spatial frequencies (20.3-5.3 cpf).

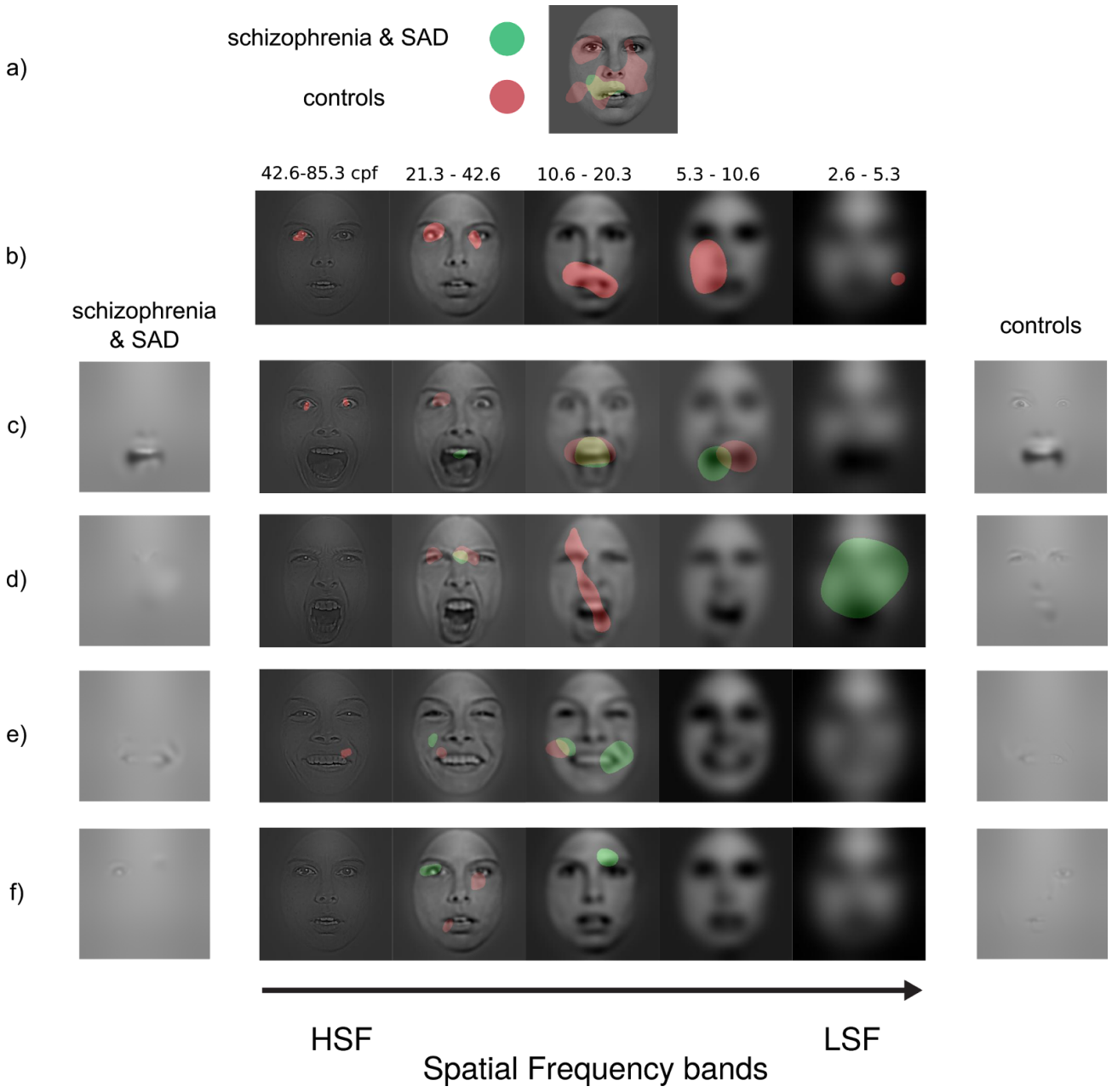


Figure 2.3. Regions attaining statistical significance ($p<.05$) in the classification images obtained during the categorization of facial expression of emotions in SZ&SAD (green) and controls (red). (a) Classification images (CI) were pooled across participants from the same subject group, scales and emotions. They show the overall use of facial features during this task for SZ&SAD

and controls. (b) CI were pooled across participants from the same subject group and emotions. They show the overall use of facial features at different scales during this task for SZ&SAD and controls. (c)-(f) CI were pooled across participants from the same subject group. They show the use of features at different scales for (c) fear, (d) anger, (e) happy and (f) neutral faces for SZ&SAD and controls. Facial features used for each emotion are also displayed on the left (SZ&SAD) and right side (controls) to help interpretation.

Discussion

A better understanding of the mechanisms underlying emotion recognition in individuals with comorbid schizophrenia and social anxiety disorder (SZ&SAD) is important considering both the prevalence and serious consequences of this additional diagnostic in schizophrenia. Here, we used a data-driven psychophysical approach and revealed, for the first time, the specific visual representations supporting facial expression of emotion recognition in SZ&SAD. We found, first, that SZ&SAD individuals have impaired recognition of facial expressions: They require more visual information and are slower than controls to recognize emotions at a 75% correct rate. This confirms and extends the evidence for emotion processing impairments in schizophrenia (SZ) patients [1] to SZ&SAD patients. Second, we showed that more severe negative symptoms (blunted affect, apathy, and reduced social drive) were associated with poorer recognition of emotions expressed facially. This also aligns with previous findings on the effects of negative symptoms in SZ [32,53–56] and generalizes them to SZ individuals with SAD. Then, we discovered that more severe negative symptoms specifically impact emotion perception through a bias for neutral facial-affect. Interestingly, affect symptoms (anxiety, depression, guilt, somatic concern) were associated with a qualitatively different confusion pattern, i.e. with higher confusions of angry faces. Considering that impaired emotion perception is especially present in individuals with SZ suffering from flat affect [53], our result of face representation biases toward neutral affect suggest that the visual representations of emotions might not only be central to perception, but also to the

expression of facial-affect in these individuals.

Another finding linking clinical psychopathology to emotion processing in individuals with SZ&SAD was that higher social anxiety (specifically fear and avoidance) symptoms actually *enhanced* the speed at which they react to facial-expressions of neutral and anger. This finding supports the idea of an important role of emotion valence in SZ&SAD's emotion processing mechanisms. Socially-phobic individuals have indeed been described to have an abnormal—but not necessarily impaired—processing of negative emotions from social stimuli [26]. Faster processing of emotions of negative valence in these individuals is generally thought to arise from an early hypervigilance to threatening social stimuli [26,57]. While it is tempting to dissociate the respective effect of social anxiety (hypervigilance to negative emotions) and negative symptoms (poor representations biased toward neutral affect) in SZ&SAD, additional experimental studies will be needed to shed light on their respective effect on facial-affect processing mechanisms in schizophrenia.

Most importantly, we showed that these emotion recognition impairments in SZ&SAD individuals could be explained by their inefficient visual representations of facial expression of emotions. Compared to typical controls, SZ&SAD's facial-affect representation relied less on fine facial cues (high spatial frequencies) and more on coarse facial cues (low spatial frequencies). Several studies with SZ individuals have found disruption in the early stage (magnocellular) process of the visual system, with, for example, impaired contrast sensitivity for low spatial frequency gratings (coarse information [27,29,58]). It has been proposed, as a consequence of this low-level processing, that SZ individuals should rely on high spatial frequencies more than controls for resolving visual tasks [39]. We clearly demonstrated that this is not the case. Instead, our results parallel and extend coarse information biases observed in two other studies having examined the role of coarse vs fine information on the classification of objects [59] and faces [60]. While these researchers concluded that these low spatial frequency biases were in opposition to

the magnocellular pathway deficits hypothesis of SZ, however, we do not agree with this interpretation. We rather believe that the impaired recognition of facial expression of emotions in SZ results from an interaction between these lower and higher-level processing mechanisms. In other words, we propose that it is the joint effect of i) their poor sensitivity to low spatial frequencies in early stages of the visual system *and* ii) the low spatial frequency bias of their higher-level (facial-affect) visual representations that produce these impairments.

Another noteworthy finding is that SZ&SAD individuals never relied on the eye regions (only on the mouth area) for their visual representations of facial expression of emotions. This systematic confusion of emotion cues from the eye regions might explain why anger and fear tend to be confused in individuals with SZ [15,17,18,32,61] as well as in the present study. Indeed, additional CI analysis on the facial information SZ&SAD relied on when they *confused* a face signal (i.e. fearful) with another emotion (i.e. anger) revealed that their confusion of these emotions emerged from their misinterpretation of cues from the eyes (see supp.fig.2.1). This analysis also showed that emotion cues systematically misinterpreted by SZ&SAD were mainly present for negative valence emotions, which adds to the idea of abnormal representations of emotion with negative valence in SZ&SAD. The abnormal representation of eye cues of SZ&SAD corroborates and extend what Lee et al. (2011) and Clark et al. (2013) found using a subset of our emotions. But this lack of reliance on the direct eye area in our SZ&SAD group appears even more pronounced than previous findings on visual emotion representations of SZ individuals. Lee et al., (2011) used the same *Bubbles* technique and found, similarly to results in our SZ&SAD group, that their SZ participants relied less on the eye regions in the highest-spatial frequencies. Lee et al. (2011)'s participants, however, still used the eye region in mid-to-high spatial frequencies to categorize both fearful and joyful faces. Clark et al., (2013) also found that SZ individuals relied on the eye region for anger recognition in all spatial-frequency bands. Our SZ&SAD group, however, never relied on the eye region for any of the emotions or at any scale. Such impaired eye representation

is also at odds with what was found, using a similar method, in (non-clinical) social anxiety participants [11]. It appears to be in line, however, with eye-tracking studies in *clinically diagnosed* socially anxious individuals[62]. Given that these studies are directly comparable to ours, more research on *clinically diagnosed* socially anxious individuals will be needed to draw more definitive conclusions. Furthering the idea that SZ&SAD have poor/atypical representations of facial-expressions, however, is our finding of their underreliance on fine information, which was observed neither in SZ, nor in individuals with social anxiety disorder. Thus, altogether, these findings suggest that schizophrenia *and* social anxiety disrupt visual emotion representations to a much greater extent than either schizophrenia or social anxiety diagnostics on their own. These highly impaired representations warrant a systematic assessment of social anxiety disorder comorbidity in samples of schizophrenia individuals.

Our finding that the visual representations of typical and SZ&SAD individuals differ may lead to a new kind of perceptual intervention for SZ&SAD individuals. Indeed, some of us recently devised a training procedure capable of inducing the use of specific facial information in neurotypical participants; and showed that the induction of the visual representation of the best face recognizers in individuals with intermediate face recognition abilities led to an enhancement of their face recognition abilities [63–65]. We hope that by inducing the visual representations of neurotypical individuals (e.g. eye emotion cues) in SZ&SAD with similar interventions will reduce their impairment at the recognition of facial expression of emotions and, in turn, help their social functioning.

Acknowledgements

NSERC Discovery grant (04777-2014) to F.G.; NSERC graduate scholarship to S.

Faghel-Soubeyrand.

Competing interests

The authors declare that there are no competing interests.

Author Contribution

F.G. & T.L. designed the research idea; S.F.S. & F.G. implemented the procedure; T.L., M.L., S.P., A.A-B. & M.V. recruited SZ&SAD patients; S.F.S. recruited neurotypical participants and supervised data collection from SZ&SAD and neurotypical participants; S.F.S. analysed the data; and S.F.S. & F.G wrote the paper with comments and additions from A.B., T.L., M.L, S.P., A.A-B. & M.V..

References

1. Green MF, Horan WP, Lee J. Social cognition in schizophrenia. *Nat Rev Neurosci*. 2015;16(10):620-631. doi:10.1038/nrn4005

2. Couture SM, Penn DL, Roberts DL. The functional significance of social cognition in schizophrenia: a review. *Schizophr Bull*. 2006;32 Suppl 1:S44-S63. doi:10.1093/schbul/sbl029

3. Irani F, Seligman S, Kamath V, Kohler C, Gur RC. A meta-analysis of emotion perception and functional outcomes in schizophrenia. *Schizophr Res*. 2012;137(1-3):203-211. doi:10.1016/j.schres.2012.01.023

4. Achim AM, Maziade M, Raymond E, Olivier D, Mérette C, Roy M-A. How prevalent are anxiety disorders in schizophrenia? A meta-analysis and critical review on a significant association. *Schizophr Bull*. 2011;37(4):811-821. doi:10.1093/schbul/sbp148

5. Braga RJ, Reynolds GP, Siris SG. Anxiety comorbidity in schizophrenia. *Psychiatry Res*. 2013;210(1):1-7. doi:10.1016/j.psychres.2013.07.030

6. Lysaker PH, Yanos PT, Outcalt J, Roe D. Association of stigma, self-esteem, and symptoms with concurrent and prospective assessment of social anxiety in schizophrenia. *Clin Schizophr Relat Psychoses*. 2010;4(1):41-48. doi:10.3371/CSRP.4.1.3

7. Pallanti S, Quercioli L, Hollander E. Social anxiety in outpatients with schizophrenia: a relevant cause of disability. *Am J Psychiatry*. 2004;161(1):53-58. doi:10.1176/appi.ajp.161.1.53

8. Achim AM, Ouellet R, Lavoie M-A, Vallières C, Jackson PL, Roy M-A. Impact of social anxiety on social cognition and functioning in patients with recent-onset schizophrenia spectrum disorders. *Schizophrenia Research*. 2013;145(1-3):75-81. doi:10.1016/j.schres.2013.01.012

9. Lecomte T, Théroux L, Paquin K, Potvin S, Achim A. Can Social Anxiety Impact Facial Emotion Recognition in Schizophrenia? *J Nerv Ment Dis*. 2019;207(3):140-144. doi:10.1097/NMD.0000000000000934

10. Gosselin F, Schyns PG. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res*. 2001;41(17):2261-2271. https://www.ncbi.nlm.nih.gov/pubmed/11448718.

11. Langner O, Becker ES, Rinck M. Social anxiety and anger identification: bubbles reveal differential use of facial information with low spatial frequencies. *Psychol Sci*. 2009;20(6):666-670. doi:10.1111/j.1467-9280.2009.02357.x

12. Lee J, Gosselin F, Wynn JK, Green MF. How do schizophrenia patients use visual information to decode facial emotion? *Schizophr Bull*. 2011;37(5):1001-1008. doi:10.1093/schbul/sbq006

13. Clark CM, Gosselin F, Goghari VM. Aberrant patterns of visual facial information usage in schizophrenia. *J Abnorm Psychol*. 2013;122(2):513-519. doi:10.1037/a0031944

14. Kohler CG, Turner TH, Bilker WB, et al. Facial emotion recognition in schizophrenia: intensity effects and error pattern. *Am J Psychiatry*. 2003;160(10):1768-1774. doi:10.1176/appi.ajp.160.10.1768

15. Huang J, Chan RCK, Gollan JK, et al. Perceptual bias of patients with schizophrenia in morphed facial expression. *Psychiatry Res*. 2011;185(1-2):60-65. doi:10.1016/j.psychres.2010.05.017

16. Bedwell JS, Chan CC, Cohen O, Karbi Y, Shamir E, Rassovsky Y. The magnocellular visual pathway and facial emotion misattribution errors in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;44:88-93. doi:10.1016/j.pnpbp.2013.01.015

17. Premkumar P, Cooke MA, Fannon D, et al. Misattribution bias of threat-related facial expressions is related to a longer duration of illness and poor executive function in schizophrenia and schizoaffective disorder. *Eur Psychiatry*. 2008;23(1):14-19. doi:10.1016/j.eurpsy.2007.10.004

18. Pinkham AE, Brensinger C, Kohler C, Gur RE, Gur RC. Actively paranoid patients with schizophrenia over attribute anger to neutral faces. *Schizophr Res*. 2011;125(2-3):174-178. doi:10.1016/j.schres.2010.11.006

19. Rapee RM, Heimberg RG. A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*. 1997;35(8):741-756. doi:10.1016/s0005-7967(97)00022-3

20. Spokas ME, Rodebaugh TL, Heimberg RG. Cognitive biases in social phobia. *Psychiatry*.

2007;6(5):204-210. doi:10.1016/j.mppsy.2007.02.006

21. Montagne B, Schutters S, Westenberg HGM, van Honk J, Kessels RPC, de Haan EHF. Reduced sensitivity in the recognition of anger and disgust in social anxiety disorder. *Cogn Neuropsychiatry*. 2006;11(4):389-401. doi:10.1080/13546800444000254

22. de Jong PJ, Martens S. Detection of emotional expressions in rapidly changing facial displays in high- and low-socially anxious women. *Behav Res Ther*. 2007;45(6):1285-1294. doi:10.1016/j.brat.2006.10.003

23. Hirsch CR, Clark DM. Information-processing bias in social phobia. *Clin Psychol Rev*. 2004;24(7):799-825. doi:10.1016/j.cpr.2004.07.005

24. Philippot P, Douilliez C. Social phobics do not misinterpret facial expression of emotion. *Behav Res Ther*. 2005;43(5):639-652. doi:10.1016/j.brat.2004.05.005

25. Heuer K, Lange W-G, Isaac L, Rinck M, Becker ES. Morphed emotional faces: Emotion detection and misinterpretation in social anxiety. *J Behav Ther Exp Psychiatry*. 2010;41(4):418-425. doi:10.1016/j.jbtep.2010.04.005

26. Machado-de-Sousa JP, Arrais KC, Alves NT, et al. Facial affect processing in social anxiety: tasks and stimuli. *J Neurosci Methods*. 2010;193(1):1-6. doi:10.1016/j.jneumeth.2010.08.013

27. Butler PD, Javitt DC. Early-stage visual processing deficits in schizophrenia. *Curr Opin Psychiatry*. 2005;18(2):151-157. doi:10.1097/00001504-200503000-00008

28. Obayashi C, Nakashima T, Onitsuka T, et al. Decreased spatial frequency sensitivities for processing faces in male patients with chronic schizophrenia. *Clin Neurophysiol*. 2009;120(8):1525-1533. doi:10.1016/j.clinph.2009.06.016

29. Brenner CA, Krishnan GP, Vohs JL, et al. Steady state responses: electrophysiological assessment of sensory function in schizophrenia. *Schizophr Bull*. 2009;35(6):1065-1077. https://academic.oup.com/schizophreniabulletin/article-abstract/35/6/1065/1843589.

30. Kim D-W, Shim M, Song MJ, Im C-H, Lee S-H. Early visual processing deficits in patients with schizophrenia during spatial frequency-dependent facial affect processing. *Schizophr Res*. 2015;161(2-3):314-321. doi:10.1016/j.schres.2014.12.020

31. Butler PD, Abeles IY, Weiskopf NG, et al. Sensory contributions to impaired emotion processing in schizophrenia. *Schizophr Bull*. 2009;35(6):1095-1107. doi:10.1093/schbul/sbp109

32. Norton D, McBain R, Holt DJ, Ongur D, Chen Y. Association of impaired facial affect recognition with basic facial and visual processing deficits in schizophrenia. *Biol Psychiatry*. 2009;65(12):1094-1098. doi:10.1016/j.biopsych.2009.01.026

33. Li C-SR. Impaired detection of visual motion in schizophrenia patients. *Prog Neuropsychopharmacol Biol Psychiatry*. 2002;26(5):929-934. https://www.ncbi.nlm.nih.gov/pubmed/12369268.

34. Notredame C-E, Pins D, Deneve S, Jardri R. What visual illusions teach us about schizophrenia. *Front Integr Neurosci*. 2014;8:63. doi:10.3389/fnint.2014.00063

35. Livingstone MS, Hubel DH. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J Neurosci*. 1987;7(11):3416-3468. https://www.ncbi.nlm.nih.gov/pubmed/3316524.

36. Bar M, Kassam KS, Ghuman AS, et al. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*. 2006;103(2):449-454. doi:10.1073/pnas.0507062103

37. Caplette L, Wicker B, Gosselin F. Atypical Time Course of Object Recognition in Autism Spectrum Disorder. *Scientific Reports*. 2016;6(1). doi:10.1038/srep35494

38. Tso IF, Carp J, Taylor SF, Deldin PJ. Role of visual integration in gaze perception and emotional intelligence in schizophrenia. *Schizophr Bull*. 2014;40(3):617-625. doi:10.1093/schbul/sbt058

39. Calderone DJ, Hoptman MJ, Martínez A, et al. Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cereb Cortex*. 2013;23(8):1849-1858. doi:10.1093/cercor/bhs169

40. First MB, Spitzer RL, Gibbon M, Williams JBW. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part I: Description. *Journal of Personality Disorders*. 1995;9(2):83-91. doi:10.1521/pedi.1995.9.2.83

41. Lukoff D, Liberman RP, Nuechterlein KH. Symptom monitoring in the rehabilitation of schizophrenic patients. *Schizophr Bull*. 1986;12(4):578-602. doi:10.1093/schbul/12.4.578

42. Ventura J, Green MF, Shaner A, Liberman RP. Training and quality assurance with the Brief Psychiatric Rating Scale:“ the drift busters.” *Int J Methods Psychiatr Res*. 1993. https://psycnet.apa.org/record/1994-27973-001.

43. Davidson JR, Miner CM, De Veaugh-Geiss J, Tupler LA, Colket JT, Potts NL. The Brief Social Phobia Scale: a psychometric evaluation. *Psychol Med*. 1997;27(1):161-166. doi:10.1017/s0033291796004217

44. Blais C, Roy C, Fiset D, Arguin M, Gosselin F. The eyes are not the window to basic emotions. *Neuropsychologia*. 2012;50(12):2830-2838. doi:10.1016/j.neuropsychologia.2012.08.010

45. Blais C, Fiset D, Roy C, Saumure Régimbald C, Gosselin F. Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*. 2017;17(7):1107-1119. doi:10.1037/emo0000283

46. Burt P, Adelson E. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on*

Communications. 1983;31(4):532-540. doi:10.1109/tcom.1983.1095851

47. Watson AB, Pelli DG. QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys*. 1983;33(2):113-120. doi:10.3758/bf03202828

48. Royer J, Blais C, Gosselin F, Duncan J, Fiset D. When less is more: Impact of face processing ability on recognition of visually degraded faces. *J Exp Psychol Hum Percept Perform*. 2015;41(5):1179-1183. doi:10.1037/xhp0000095

49. Tardif J, Morin Duchesne X, Cohan S, et al. Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychol Sci*. 2019;30(2):300-308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338.

50. Rutishauser U, Tudusciuc O, Wang S, Mamelak AN, Ross IB, Adolphs R. Single-Neuron Correlates of Atypical Face Processing in Autism. *Neuron*. 2013;80(4):887-899. doi:10.1016/j.neuron.2013.08.029

51. Ince RAA, Jaworska K, Gross J, et al. The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex*. 2016;26(11):4123-4135. doi:10.1093/cercor/bhw196

52. Chauvin A, Worsley KJ, Schyns PG, Arguin M, Gosselin F. Accurate statistical tests for smooth classification images. *Journal of Vision*. 2005;5(9):1. doi:10.1167/5.9.1

53. Gur RE, Kohler CG, Ragland JD, et al. Flat Affect in Schizophrenia: Relation to Emotion Processing and Neurocognitive Measures. *Schizophrenia Bulletin*. 2006;32(2):279-287. doi:10.1093/schbul/sbj041

54. Kohler CG, Bilker W, Hagendoorn M, Gur RE, Gur RC. Emotion recognition deficit in schizophrenia: association with symptomatology and cognition. *Biological Psychiatry*. 2000;48(2):127-136. doi:10.1016/s0006-3223(00)00847-7

55. Leszczyńska A. Facial emotion perception and schizophrenia symptoms. *Psychiatr Pol*. 2015;49(6):1159-1168. doi:10.12740/PP/38919

56. Daros AR, Ruocco AC, Reilly JL, Harris MSH, Sweeney JA. Facial emotion recognition in first-episode schizophrenia and bipolar disorder with psychosis. *Schizophr Res*. 2014;153(1-3):32-37. doi:10.1016/j.schres.2014.01.009

57. Aronoff J, Barclay AM, Stevenson LA. The recognition of threatening facial stimuli. *Journal of Personality and Social Psychology*. 1988;54(4):647-655. doi:10.1037/0022-3514.54.4.647

58. Martinez A, Hillyard SA, Dias EC, et al. Magnocellular Pathway Impairment in Schizophrenia: Evidence from Functional Magnetic Resonance Imaging. *Journal of Neuroscience*.

2008;28(30):7492-7500. doi:10.1523/jneurosci.1852-08.2008

59. Laprevote V, Oliva A, Thomas P, Boucart M. 373 – Face processing deficit in schizophrenia: Seeing faces in a blurred world? *Schizophrenia Research*. 2008;98:187. doi:10.1016/j.schres.2007.12.440

60. Laprevote V, Oliva A, Ternois A-S, Schwan R, Thomas P, Boucart M. Low Spatial Frequency Bias in Schizophrenia is Not Face Specific: When the Integration of Coarse and Fine Information Fails. *Front Psychol*. 2013;4:248. doi:10.3389/fpsyg.2013.00248

61. Habel U, Chechko N, Pauly K, et al. Neural correlates of emotion recognition in schizophrenia. *Schizophr Res*. 2010;122(1-3):113-123. doi:10.1016/j.schres.2010.06.009

62. Claudino RG e., de Lima LKS, de Assis EDB, Torro N. Facial expressions and eye tracking in individuals with social anxiety disorder: a systematic review. *Psicologia: Reflexão e Crítica*. 2019;32(1). doi:10.1186/s41155-019-0121-8

63. Faghel-Soubeyrand S, Dupuis-Roy N, Gosselin F. Inducing the use of right eye enhances face-sex categorization performance. *J Exp Psychol Gen*. 2019;148(10):1834-1841. doi:10.1037/xge0000542

64. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. *Nature*. 2005;433(7021):68-72. doi:10.1038/nature03086

65. DeGutis J, Cohan S, Nakayama K. Holistic face training enhances face processing in developmental prosopagnosia. *Brain*. 2014;137(6):1781-1798. doi:10.1093/brain/awu062

Chapitre 3 –Article 2

Diagnostic Features for Human Categorisation of Adult and Child Faces

Simon Faghel-Soubeyrand[1,2], Juliane A. Kloess[2], Frédéric Gosselin[1], Ian Charest[1,2], and Jessica Woodhams[2]

[1]Département de Psychologie, Université de Montréal

[2] School of Psychology, University of Birmingham

Author Note

Simon Faghel-Soubeyrand https://orcid.org/0000-0002-7867-8589
Juliane A. Kloess https://orcid.org/0000-0002-8342-7880
Frédéric Gosselin https://orcid.org/0000-0002-3797-4744
Ian Charest https://orcid.org/0000-0002-3939-3003
Jessica Woodhams https://orcid.org/0000-0002-9674-6653

The authors have no known conflict of interest to disclose.

The authors would like to express their gratitude and appreciation to West Midlands Police for their assistance, time and effort in supporting the study undertaken and presented here, as well as the Institute for Global Innovation for financially supporting this study.

Correspondence concerning this article should be addressed to Jessica Woodhams, School of Psychology, University of Birmingham, 52 Pritchatts Road, Edgbaston, Birmingham, B15 2SA, UK. E-mail: J.Woodhams@bham.ac.uk

Publié : Faghel-Soubeyrand S, Kloess JA, Gosselin F, Charest I and Woodhams J (2021) Diagnostic Features for Human Categorisation of Adult and Child Faces. *Front. Psychol.* 12:775338. doi: 10.3389/fpsyg.2021.77533

Abstract

Knowing how humans differentiate children from adults has useful implications in many areas of both forensic and cognitive psychology. Yet, how we extract age from faces has been surprisingly underexplored in both disciplines. Here, we used a novel data-driven experimental technique to objectively measure the facial features human observers use to categorise child and adult faces. Relying on more than 35,000 trials, we used a reverse correlation technique that enabled us to reveal how specific features which are known to be important in face-perception—position, spatial-frequency, and orientation—are associated with accurate child and adult discrimination. This showed that human observers relied on evidence in the nasal bone and eyebrow area for accurate adult categorisation, while they relied on the eye and jawline area to accurately categorise child faces. For orientation structure, only facial information of vertical orientation was linked to face-adult categorisation, while features of horizontal and, to a lesser extent oblique orientations, were more diagnostic of a child face. Finally, we found that spatial-frequency diagnosticity showed a U-shaped pattern for face-age categorisation, with information in low and high spatial frequencies being diagnostic of child faces, and mid spatial frequencies being diagnostic of adult faces. Through this first characterisation of the facial features of face-age categorisation, we show that important information found in psychophysical studies of face-perception in general (i.e. the eye area, horizontals, and mid-level spatial-frequencies) is crucial to the practical context of face-age categorisation, and present data-driven procedures through which face-age classification training could be implemented for real-world challenges.

Public significance statement

Understanding how we recognise children from adults has wide-reaching applications, such as the improvement of the classification of indecent images of children by police analysts and computer vision algorithms. Here we show a thorough characterisation of the visual strategies human observers use to recognise children from adults, providing data-driven associations between specific facial features and human accurate age classification.

Introduction

The amount of imagery depicting child sexual abuse (referred to here as indecent images of children, IIOC[2]; Edwards, 2013) in circulation has dramatically increased in the last 25 years, from estimates of thousands of such images in the late 1990s to millions or tens of millions nowadays (Home Office, 2017; Krasodomski-Jones et al., 2017). In the United Kingdom, the task of assessing whether a suspect's digital material includes IIOC is conducted by digital forensics analysts who are members of specialist teams in police forces across the UK. This task involves determining whether a given image a) depicts a child, and b) if it is of an indecent nature (Kloess et al., 2019). In other countries, professionals undertaking the same task have the added challenge of distinguishing between children within specific age bands (e.g., in Germany, distinctions are made regarding IIOC that depict children under 18 years and children under 14 years; (Ratnayake et al., 2014).

Identifying and classifying IIOC is a time-consuming task which is also stressful and emotionally challenging (Franqueira et al., 2018; Ratnayake et al., 2014). In addition, some imagery presents particular difficulties for the human classifier (Kloess et al., 2019; Ratnayake et al., 2014). Unsurprisingly, the potential to (semi-)automate the identification and classification of IIOC using algorithms is currently being explored (e.g., Gao & Ai, 2009; Home Office, 2017; Sae-Bae et al., 2014). Developing our understanding of what makes children's faces distinct from adult faces, in particular the features used by human observers to discriminate them, is therefore an important step in developing software and advancing training for digital forensics analysts (and allied professionals). This study builds on existing qualitative research that has explored the aspects and attributes within imagery that digital forensics analysts report drawing on in order to inform their decision-making in the process of identifying and classifying IIOC, including specific facial and bodily features of children (Kloess et al., 2019, 2021; Michalski et al., 2019), by applying a new (data-driven) reverse-correlation technique relying on Gabor wavelets to this

[2] This material is also referred to as child sexual exploitation material (Merdian et al., 2021), child pornography and sexually exploitative material of children (Franqueira et al., 2018).

problem area. In doing so, it fills important gaps in the literature, given i) the scarcity of studies that have examined facial differences associated with age (Gao & Ai, 2009) and, most importantly, ii) the absence of studies having revealed the specific facial features human observers use for age categorisation.

There are many physical changes associated with the maturation of faces. During childhood, craniofacial growth alters the shape and size of a child's face, as well as the relative positions of different facial features. These changes are due to the development of permanent teeth and puberty. The greatest changes are seen between birth and the age of five years, with the rate of change being non-linear (Ricanek et al., 2015). Compared to adult faces, children's faces are typically characterized by a "protruding forehead, large head, round face, big eyes and a small nose or mouth" (Komori & Nittono, 2013, p. 285).

Findings from research on the border-control task of the facial matching of children demonstrate that differentiating between faces of similar-aged children is difficult for non-expert (Kramer et al., 2018) and expert (i.e., nursery workers and super-recognizers; (Bate et al., 2020; Belanova et al., 2018) observers, and that this task is harder the younger the child who is depicted (Michalski et al., 2019). In combination with research regarding how child and adult faces differ physically, this suggests that, particularly with younger children, their faces are defined by a common set of qualities which differ to the faces of adults. While these qualities may make the task of face-matching children more difficult, they should make the task of discriminating between child and adult faces easier, assuming that those who are engaged in the discrimination task attend to valid features for age categorisation.

In relation to this, (Kloess et al., 2019) assessed the inter-rater agreement of law enforcement personnel experienced at identifying and classifying IIOC, finding levels of agreement that were not always adequate. In subsequent focus groups, the officers reported that features they found more indicative of a younger age included large eyes relative to the rest of the

face, the presence of milk teeth, or the eruption of adult teeth[3]. They also reported that the classification task was more challenging when there was a mismatch between the apparent maturity of the depicted person's face and their body (or vice-versa), and where the depicted person was made to look younger or older (e.g., by means of make-up and clothing). More recently, (Kloess et al., 2021) followed up on their 2019 study by asking digital forensics analysts to describe the key facial attributes they use in determining the presence of a child in imagery. These included: (i) large eyes, (ii) a small nose, (iii) a round facial shape, (iv) an absence of cheekbones, (v) early signs of teeth development, and (vi) smooth skin.

Attending to the face (rather than other cues) may thus improve the accuracy of the identification and classification of IIOC. It is therefore important to further investigate (i) whether participants do indeed rely on the features previously identified by police staff when discriminating between child and adult faces, and, relatedly, (ii) whether attending to these features/facial areas does indeed lead to more accurate decision-making. In this matter, psychophysical studies on face perception have outlined the behavioural relevance of important features of facial images, such as orientation structure (Duncan et al., 2017; Valerie Goffaux et al., 2016; Valerie Goffaux & Greenwood, 2016), spatial location (Adolphs et al., 2005; Faghel-Soubeyrand et al., 2019; Gosselin & Schyns, 2001; Tardif et al., 2019), colour (Benitez-Quiroz et al., 2018; Dupuis-Roy et al., 2019; Young et al., 2013), and spatial frequency (Estéphan et al., 2018; Faghel-Soubeyrand et al., 2020; Tardif et al., 2017). However, while the use of these physical properties have been revealed for various face-processing tasks (e.g. face-identification, face-expression & face-gender recognition in the above-cited studies), the diagnostic features for face-age recognition have never been explored.

The present study tackles this question by employing a recently developed reverse-correlation technique, called Diagnostic Feature Mapping (DFM; (Alink & Charest, 2020), in order to precisely quantify the use of three core physical properties of facial images of

[3] The use of dentition developmental stage in making accurate age assessments has been questioned, and may therefore only be useful in images where the depicted person's mouth is open (see Ratnayake et al., 2014), which is not always the case in the context of IIOC.

children and adults, namely location, orientation structure and spatial frequency. These properties are randomly sampled in a similar way to Bubbles (Gosselin & Schyns, 2001). A multiple linear regression applied to these samples and response accuracy can reveal the diagnosticity of behaviourally relevant features of images which are known to be crucial to the human visual system (i.e., different spatial-frequencies, orientations, spatial coordinates and sizes in the striate cortex; (Hubel & Wiesel, 1968). Contrasting with previous psychophysical studies, however, is the capacity of DFM to reveal these three core features in a single experimental paradigm, permitting a thorough description of the facial information relied for categorization tasks. Using such Gabor features was also motivated by studies in computer vision (Gao & Ai, 2009), suggesting that models using Gabors as priors perform well in explaining differences between facial images of children and adults. DFM therefore provided an informed, comprehensive and entirely data-driven way to reveal the specific facial features associated with face-age categorisation.

Method

Participants

A screening questionnaire was used to screen for conditions that would prevent a volunteer being able to take part in the experiment. This resulted in one volunteer being excluded. Of the 16 participants who took part in the experiment (male = 9, female = 7, right-handed = 14, left-handed = 2, White British = 12, and Asian-Indian = 4, aged between 21-47 years (M = 38.44, SD = 8.33), seven were students or staff at the University of Birmingham, and nine were digital forensics analysts recruited from West Midlands Police. All participants had no history of psychiatric diagnosis and all had normal or corrected-to-normal vision. Full ethical approval for the study was granted by the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham (ERN_15-1374AP10A). In addition, the researchers adhered to the British Psychological Society's Code of Ethics and Conduct (2018) throughout the study, and the experiment was conducted in accordance with the Declaration of Helsinki.

Participants were recruited through the University of Birmingham's Research Participation Scheme, via the last author's professional network at the University of Birmingham, and by means of an advertisement that was circulated via internal channels within West Midlands Police to officers who were working with child sexual exploitation materials on a daily basis. This study was not preregistered.

Stimuli

A total of 12 images depicting faces of either children or adults (6 adults, 6 males; all with neutral expressions) were selected from the Radboud Face dataset (Langner et al., 2010), and converted into 250 x 250 pixels grayscale images. Each image was aligned based on twenty landmarks (averaged to six mean coordinates for left and right eyes, left and right eyebrows, nose and mouth) using Procrustes transformations (rotation, scaling and translation), and revealed through an elliptical mask that excluded non-facial cues such as hair, ears and neck. The luminance profile of the resulting face images was equalized across images using the SHINE toolbox (Willenbockel et al., 2010).

Gabors Wavelet Decomposition

A more in-depth description of the Gabors wavelet decomposition algorithm has been published elsewhere (Alink & Charest, 2020). Briefly, we used a custom-made Matlab program which aimed to reduce each face image to a subset of the most important 2200 Gabor wavelets. To do so, we considered wavelets of 20 different spatial frequencies (SFs), exponentially increasing between 2.4 and 87 cycles per image ($SF = \frac{10}{45} \times 1.08^{n}; n = [1, 2, ..., 20]$). They had 12 orientations between 15° to 180° in constant steps of 15°, and were centred on each possible pixel of the input grayscale image. The final sets of Gabor features were selected based on their best fit (least-square correlation) to the original grayscale images. Amplitudes were set to an equal value for all wavelets. Partial reconstructions of the images at every given trial were created by randomly selecting a subset of the 2200 Gabor wavelets, and summating them (see Figure 3.1a). The resulting range of pixel values was modified linearly so that all stimuli covered the full 0-255 range.

Apparatus

The experimental program ran on Windows computers in the MATLAB environment (The Mathworks, Natick, MA), using functions from the Psychophysics Toolbox (Brainard, 1997; Watson & Pelli, 1983) 1997). Stimuli were shown on 22-in. monitors (1920 x 1080 pixels at 60Hz). Participants performed the experiment in a dimly lit room, and viewing distance was maintained at ~76 cm.

Experimental procedure

The experiment took place in the laboratories of the School of Psychology at the University of Birmingham. Participants were first given standardized instructions about the experimental task. Each trial began with a grey screen and a central fixation point (250 ms), which participants were instructed to gaze at. The partial face reconstruction spanning 8 deg of visual angle was then shown in the central area of the screen, and remained visible until the participant's response. Participants were asked to categorise the faces depicted in the stimuli as either children or adults as fast and as accurately as possible. The stimulus presentation order was randomized for each participant. All 16 participants completed 2400 trials of this adult vs. child face discrimination task, with short breaks every 200 trials (approximately 10 minutes). The task was completed in one session that lasted between 1.5-2 hours. Overall, a total of 38,400 trials were recorded for this study. Estimates of effect size could not be assessed considering the lack of previous studies on face-age recognition, but the current number of observations per participant is more than twice the typical range used in reverse-correlation studies (e.g. Tardif et al., 2019).

The quantity of face information (i.e., the number of Gabor Wavelet features) necessary to maintain an accuracy of 75% was adjusted on a trial-by-trial basis with the QUEST algorithm (Watson & Pelli, 1983). Royer and colleagues (Royer et al., 2015) showed that such a threshold (measured using a similar method) is strongly correlated with three commonly used face recognition ability tests (r=-.79 with the mean of the Cambridge Face Memory Test +, the Cambridge Face Perception Test, and the Glasgow Face Matching Test short version). We thus

used the number of Gabors required to maintain an accuracy of 75% as our individual index of face-age discrimination performance.

After the experiment, participants were verbally debriefed and were given a debriefing sheet to take with them. Participants recruited through the Research Participation Scheme were compensated with course credits and participants from West Midlands Police were reimbursed for their travel expenses and time.

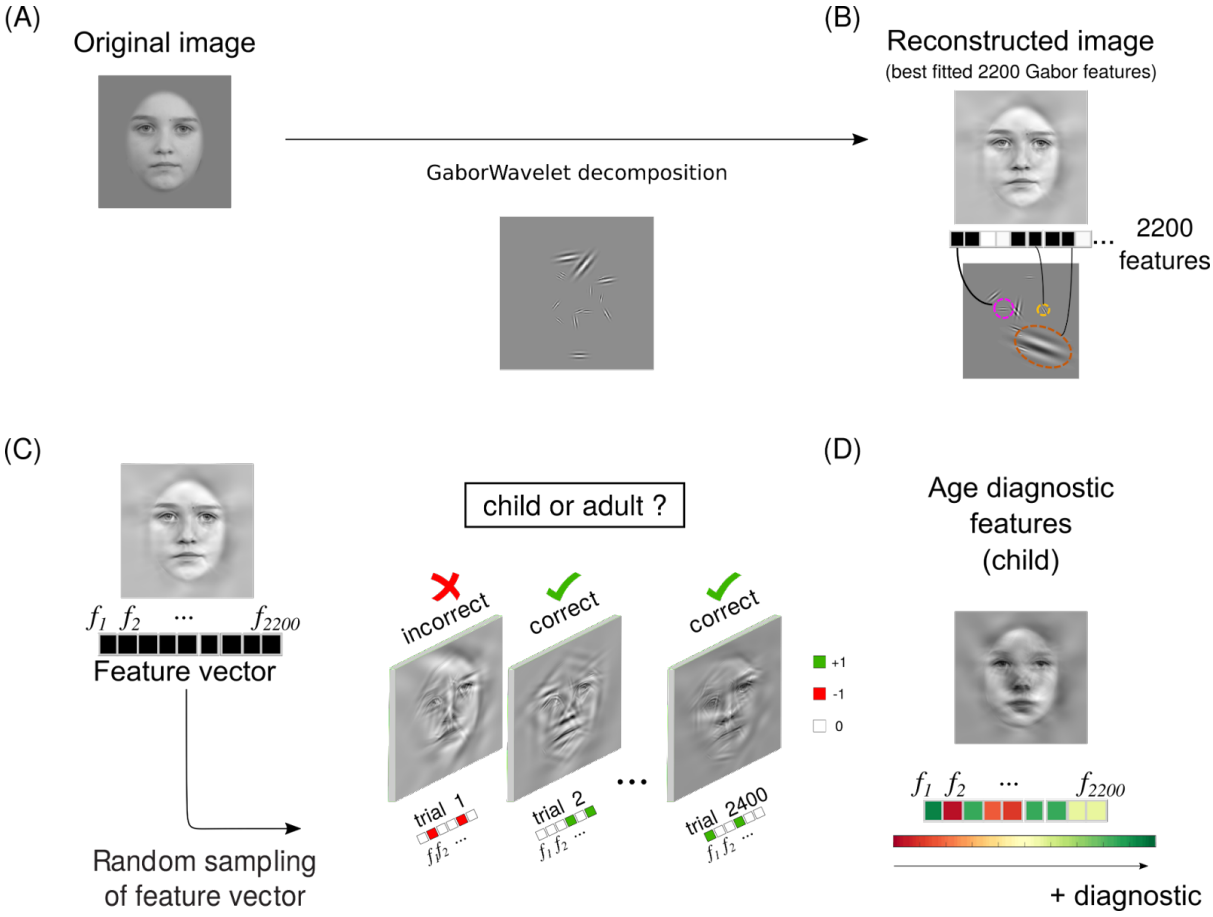


Figure 3.1. Image feature decomposition and reverse-correlation procedure. A reconstructed image (B, lower image) was created using a subset of 2200 Gabors wavelets varying in spatial-frequency (from 2.4 to 87 cpi), orientation (10° to 180°), x-y coordinates and size. These 2200 wavelets, the most correlated with the original image (A), are referred to as Gabor features (see also (Alink & Charest, 2020) for further details). (C) At any given trial, the stimulus was created by randomly sampling the 2200 Gabor features to reveal parts of the reconstructed image. Participants were asked to indicate whether the resulting sampled face was a child or an adult. Each participant was submitted to 2.4K trials, for a grand total of >38k trials. (D) We obtained

the diagnostic features by weighing and summing the randomly sampled feature vectors across trials with z-scored behavioural accuracies.

Data and code availability

The data and MATLAB code that support the findings of this study are available from the corresponding author upon reasonable request.

Feature Diagnosticity Index

We quantified how the Gabor features (the Gabor wavelets) were differently associated with face representations by computing a Feature Diagnosticity index (FDi; see (Alink & Charest, 2020) for each of the features composing our face images, for every participant, and stimulus independently. This was done by weighing the features presented across trials by the z-scored accuracy values, providing us with a three-dimensional matrix of FDi with the dimensions number of participants (n = 16), number of images (12) and number of features (2200).

Results

Accuracy was close to the target 75% (M_{acc} = .7577, SD_{acc}= .0443). One participant was excluded from further analysis because of poor performance (number of Gabors threshold was greater than 3 standard deviations above the mean of all participants). The average number of Gabors required to attain 75% accuracy ranged from a low of 156.52 (i.e. best performance) to a high of 443.75 (i.e. lowest performance; $M_{nbgabors}$=275.52, $SD_{nbgabors}$=73.86).

Facial areas for child and adult classification

To reveal the facial attributes used to categorise faces as child vs. adult, we reprojected each Gabors FDI to their specific x-y image position, smoothed (with a Gaussian kernel with a standard deviation of 10 pixels) the resulting image regression coefficients for each stimulus, and compared them between children and adults faces with unpaired t-tests. These comparisons of regression coefficients are displayed on Figure 3.2a, with hot colors (positive t-statistics)

indicating evidence for accurate adult classification, and cold colors (negative values) indicating child classification evidence. Significant facial features (*ps* < .05) for adult classification evidence were found in an area around the nasal bone and eyebrows structure. In contrast, we found that the eyes, nose, and jawline were used as evidence for child face categorisation. The significant regions have been highlighted on an average of all face stimuli on Figure 3.2b for child evidence (left bottom image) and adult evidence (right bottom image). The average regression coefficient image (irrespective of category type) is shown in supplementary Figure 3.1A.

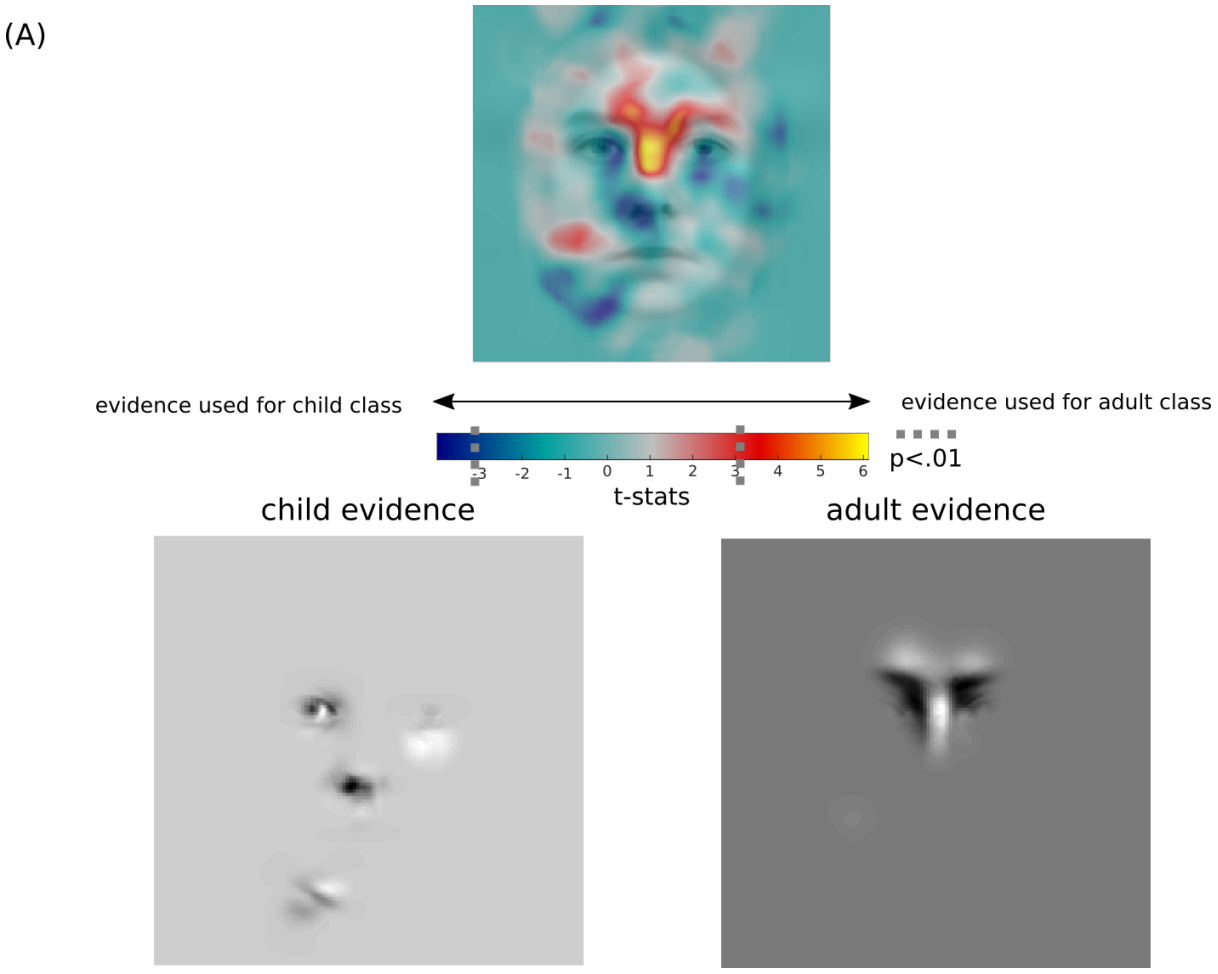


Figure 3.2. Facial features for child and adult categorisation. Regression coefficients for 2D feature positions (for accurate classification of child and adult images) were spatially smoothed,

and compared between child and adult faces with unpaired t-tests. On the first panel, hot colors (positive t-statistics) indicate evidence for accurate adult classification, and cold colors (negative values) indicate evidence for accurate child classification. The two bottom panels show pixels (i.e., regions of faces) significantly linked to accurate child and adult categorisation, respectively.

Spatial-frequencies for child and adult classification

To determine the spatial frequencies (or information granularity) used to accurately categorise child and adult faces, we correlated the spatial-frequency parameters of the presented Gabor features with accurate responses for child and adult faces across trials for all participants. The regression coefficients were compared between child and adult faces with unpaired t-tests. The resulting evidence magnitude for the 20 SFs bands are shown in Figure 3.3, with significant SF bands (t-stats with $p<.05$) highlighted. Overall, this revealed that low (coarse details of 5.1 cpi) and high SFs (fine details of 49.5 cpi) were used as evidence of child faces, while mid-SFs were relied on by our participants as evidence for adult classification. The average SFs regression coefficient (irrespective of category type) are shown in supplementary Figure 3.1B.

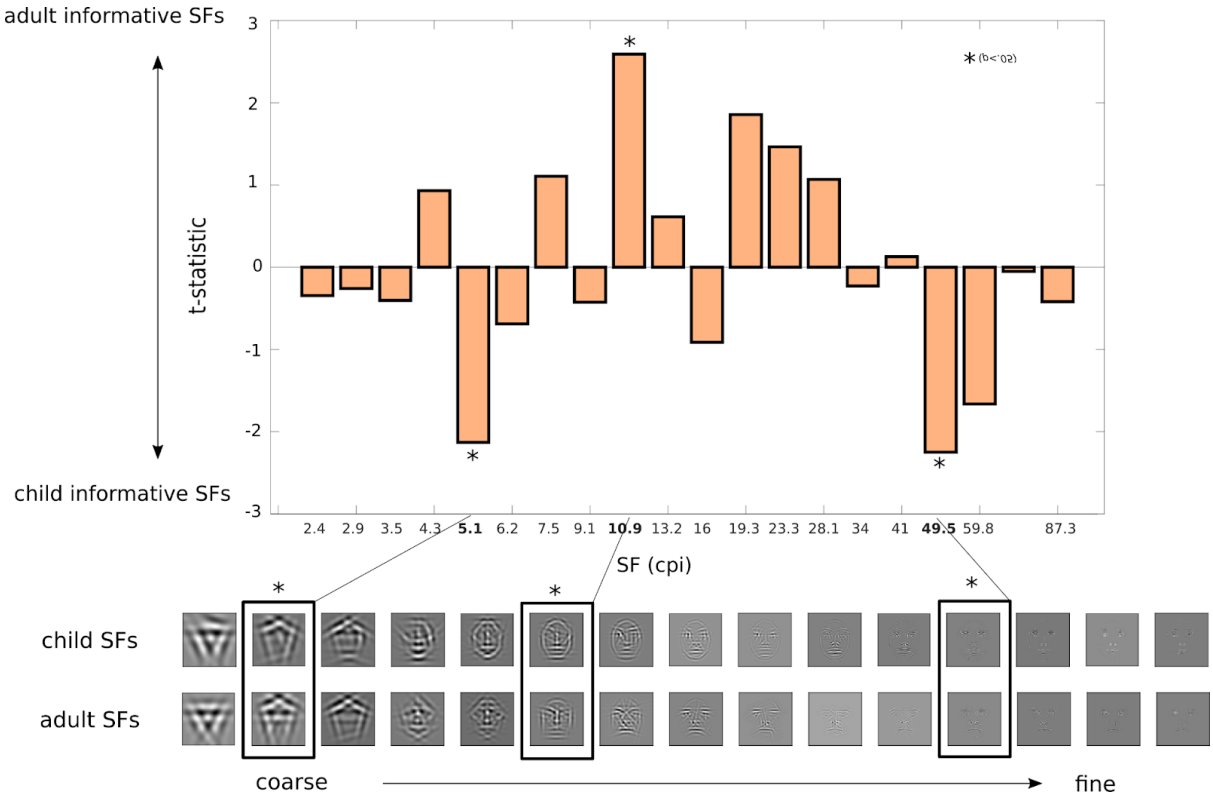


Figure 3.3. Spatial frequency evidence for categorisation of child and adult faces. Correlation coefficients for feature granularity (for accurate classification of child and adult images) were compared between child and adult faces for each one of the 20 SF bins (2.4-87.3 cpi) with unpaired t-tests. For visualisation purposes, the lower panel images display the facial information for each of the SF bands for child and adult faces. Here, positive t-statistics shows SF evidence for child-like faces, while negative values shows SF evidence for adult-like faces. Significant SF bands —low SF of 5 cpi and high SF of 50 cpi for child faces, mid-SF of 22 cpi for adult faces—are outlined with an asterisk.

Orientation of facial traits for child and adult classification

To determine the type of orientation structure relied upon for accurate categorisation of child and adult faces, we correlated the orientation parameter of the presented Gabor features with accurate responses for child and adult faces across trials for all participants. The regression coefficients were then compared between child and adult faces with unpaired t-tests. The corresponding t-statistics are shown in Figure 3.4, with significant orientation bands (t-stats with

p<.05) highlighted. Overall, this revealed that horizontal (90°) and oblique facial structures (60° and 150°) were used as evidence for classification of a face as a child. For adult face categorisation, however, vertical features (180°) were linked to accurate responses in our participants. The average orientation regression coefficient (irrespective of category type) is shown in supplementary Figure 3.1A.

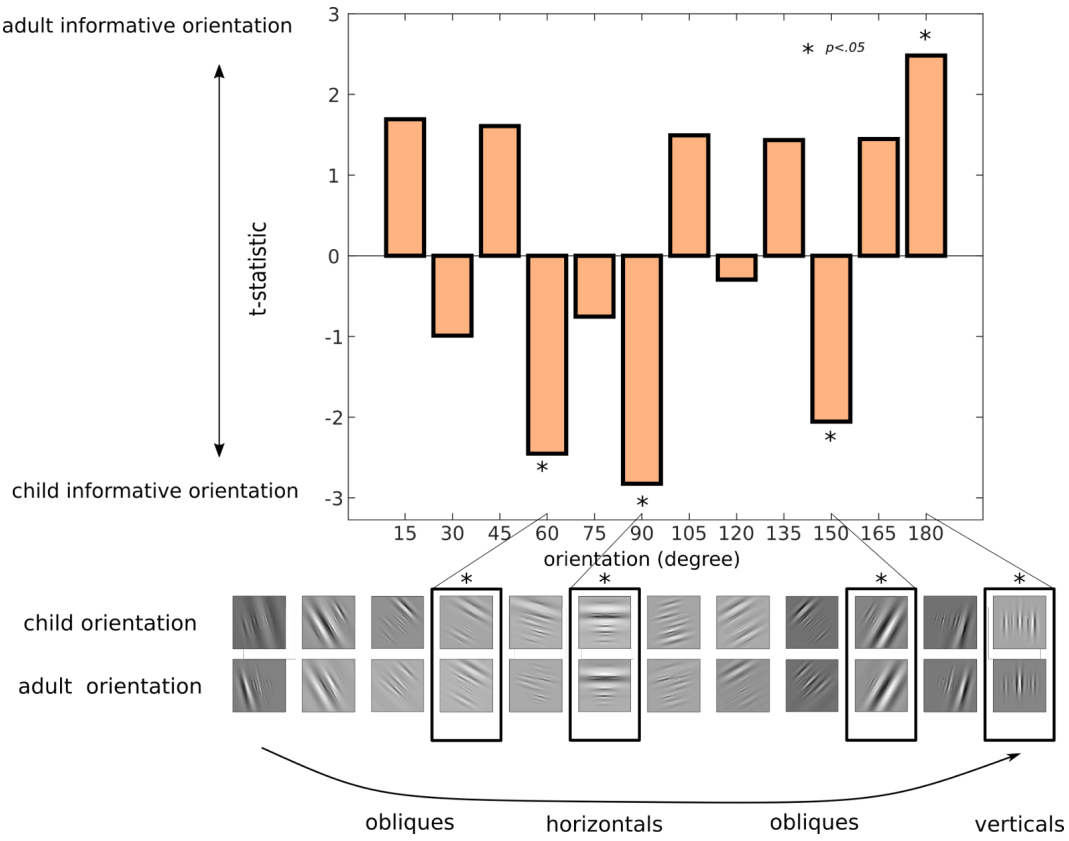


Figure 3.4. What orientation structure do we rely on to categorise child and adult faces? Correlations coefficients for feature orientation (for accurate classification of child and adult images) were compared between child and adult faces for each one of the 12 orientation bands (10° to 180° in 15° steps) with unpaired t-tests. For visualization purposes, the lower panel images display the facial information for each of the 12 orientation bands for child and adult faces. Positive t-statistics shows orientation evidence for child-like faces, while negative values shows orientation evidence for adult-like faces. Significant orientation bands —horizontal (90°) and oblique (60° & 150°) information for child faces, vertical information (180°) for adult faces—are outlined with an asterisk.

Discussion

What specific visual features drive our accurate categorisation of faces as either a child or an adult? One practical implication of answering this question is the improvement of the identification and classification of indecent images of children (IIOC; Edwards, 2013) which involves determining whether a given image a) depicts a child, and b) if it is of an indecent nature (Home Office, 2017; Kloess et al., 2019). Here, we used a data-driven approach—Diagnostic Feature Mapping (DFM; Alink & Charest, 2020; see also Gosselin & Schyns, 2001 for a similar approach) to reveal the specific facial features human participants draw on in order to distinguish child vs. adult faces. Specifically, we revealed the use of three types of features important to the visual system—spatial-frequency, position, and orientation—that human observers extract while they complete child vs. adult categorisations, yielding a precise and comprehensive characterization of the relevant facial information for face-age categorisation.

The participants in the present study used information located around the nasal bone and eyebrows (supraorbital ridge) to classify adult faces. There was some evidence that they also used an area of the cheek, which may correspond with the flattening of the cheeks that occurs during the transition from childhood to adulthood (Kau & Richmond, 2008). In contrast, the area around the eyes, nose and jawline were more indicative of a child's face. Indeed, this matches with physical differences in faces of children: large eyes relative to the rest of the face, a small, wider, up-turned nose with a concave bridge, and a rounder chin, are all physiological characteristics that have been reported (Liggett, 1975; Sforza et al., 2011). Our findings are also in line with some of the facial features reported by digital forensics analysts in Kloess et al. (2021), who explicitly described the following facial attributes as indicative of a younger age: (i) large eyes, (ii) a small nose, (iii) a round facial shape, and (iv) an absence of cheekbones. The present study corroborated these findings by revealing, with implicit and data-driven methods, the facial attributes linked to accurate child vs. adult categorisation. But our DFM paradigm enabled us to go further than revealing facial areas for accurate face-age categorisation from faces.

Indeed, DFM's reliance on Gabor wavelets parameters enabled us to show the specific level of detail (spatial-frequencies, SF) and orientation structure humans extract from images of faces in order to determine the age from a face. We found that the spatial-frequencies associated with age classification manifested a U pattern from low to high SF, with low SF (coarse features) and high SF (fine features) being associated with child faces, and mid-level SF being associated with adult faces. The fact that coarse facial features are associated with younger, child-like faces aligns with the layman notion that some facial features of children are larger, such as the forehead, the head, and the eyes (Komori & Nittono, 2013). Interestingly, the mid-level SFs (10 cpi) found to be indicative of adult faces matches the level of information required for (adult) facial identification found in previous studies (e.g., (Tardif et al., 2017, 2019; Willenbockel et al., 2010). This might explain, in turn, why identification of children's faces has been reported to be so difficult in lab conditions (Bate et al., 2020; Belanova et al., 2018; Kramer et al., 2018), as well as in real-world conditions (e.g. Kloess et al., 2019, 2021); the level of detail of facial information for our classification of a face as a child (i.e. low & high-SFs) does not match the facial information we normally use for adult face-identification (i.e. mid-SFs).

For orientation information, we found that vertical features of a face, presumably around the nasal bone, is indicative of an adult's face. Indeed, as shown in Figure 3.2, the eyebrow region as well as the nasal bone were associated with accurate identification of stimuli as an adult. These features coincide with areas of the skull that reliably differentiate the sexes in adult skulls, namely the glabella and the supraorbital ridge (Graw et al., 1999; Tanner & Tanner, 1990). Between puberty and adulthood, men develop a prominent supraorbital ridge and a larger glabella; in younger males, these features are not as apparent leading to uncertainty when classifying younger skulls as male or female (Graw et al., 1999). Our participants may therefore be using features reliably associated with adult male faces to differentiate child from adult faces in our sample.

In contrast, oblique features (i.e., specifically information at angles of 60° and 150°) were more indicative of a child's face, presumably from the rounder jawline and eyes of children's

faces. In addition, we found that horizontal information—which is paramount for face identification and the categorization of facial expressions of emotions (Duncan et al., 2017; Valerie Goffaux et al., 2016; Valérie Goffaux & Dakin, 2010; Valerie Goffaux & Greenwood, 2016; Pachai et al., 2013)—was significantly associated with child categorisation. In fact, information for face-age categorisation peaked around this horizontal band (specifically at 75° when averaged across both child and adults). This finding is yet another piece of evidence showing that horizontal information is at the core of face processing tasks, and extends the importance of this information for the first time to the challenging task of face-age categorisation. Our findings also go a step further than previous psychophysical studies of face perception by showing that it is the horizontal information in the eye region that is the most relevant in the context of this face processing task. To our knowledge, this is the first time horizontal information is linked in a direct manner to eye information in a face processing task (e.g., see Duncan et al., 2017 for an indirect link).

In terms of the real-world implications of our findings, there appears to be a common set of facial features that is associated with children's faces, which may explain why differentiating between faces of different children can present such a challenge (e.g., the task facing border control agents; (Bate et al., 2020; Belanova et al., 2018; Kramer et al., 2018; Michalski et al., 2019). Our findings are also relevant to the task of identifying IIOC, whether by a human or an AI algorithm, and provides an evidence base for the training of digital forensics analysts (and applied professionals) who have to differentiate between faces of children from those of adults. It is reassuring that our findings coincide with some of the features described in accounts given by experts in this task as to the features to which they attend (Kloess et al., 2021). However, by incorporating our findings into training, this knowledge can be imparted to new recruits, ensuring that they are using the features empirically demonstrated to be associated with accurate differentiation from the start. Our findings are also relevant to discussions about verifying children online for safeguarding (GCHQ, 2020), and to technology companies who design age assurance algorithms that interface with a webcam (e.g., Yoti).

Finally, it is important to acknowledge the limitations of our study. The facial stimuli represented 12 individuals and all were of a White ethnicity. Three key implications stem from this: (i) our method needs replication with a larger set of faces; (ii) it should be extended to the study of faces of other ethnicities (Gao & Ai, 2009); and (iii) the interaction between the ethnicity of the viewer and the ethnicity of the stimuli should be evaluated. According to the Global Threat Assessment completed by the WeProtect Global Alliance (2019), children from lower income countries are at particular risk of sexual exploitation and abuse in order to provide family income given the high level of poverty in some of these areas. This also results in an increase of non-White children being depicted in IIOC, which the analysts in the Kloess et al. (2021) study reported as presenting an additional challenge due to the lack of familiarity with those of a different ethnicity to their own, and the fact that different ethnicities can represent great variation in terms of the stages and nature of physical development.

In summary, we used a comprehensive data-driven technique to reveal how three major types of facial information are used by human observers to make face-age categorisations. Notably, the sampling of all three visual features—position, spatial-frequency, orientation—simultaneously in the DFM procedure permitted us to interpret and integrate these findings in a single experiment/data set, relating spatial information (e.g., features of the jawline) with orientation information (e.g., oblique information). The features of the face our participants relied on to differentiate child from adult faces align with what is known about changes in facial physiology from childhood to adulthood, and with some of what digital forensics analysts have self-reported about the features to which they attend when trying to classify IIOC. These results are the first to show the *relevance* of specific facial features for age classification by providing evidence that attending to these features (over others) is associated with *accurate* face-age classification.

In this study, we employed reverse correlation to reveal the key facial features used by human observers for face-age categorisation. The diagnostic image features that we identified correspond to core psychophysical features of human face perception (i.e., features in the eyebrow and nose area, those in the mid spatial-frequencies, and those around the horizontals). This not only advances our knowledge of the psychophysical correlates for face-age categorisation for the first time, but also constitutes a crucial step towards implementing evidence-based face-age classification training. Indeed, the features we identified could be used to inform current efforts to identify child versus adult faces, be this by human or machine (e.g., identifying and classifying IIOC or for age verification purposes). We have already shown, for example, that inducing the use of the best set of features in human observers is fruitful in enhancing their performance for other face classification tasks (Faghel-Soubeyrand et al., 2019). Police-analysts could thus be trained to focus their attention on the features outlined here to enhance their performance in detecting a child in IIOC. Another interesting application of our findings would be to build better automatic (computer) algorithms for child vs. adult face classification by biasing the weights of pattern recognition algorithms (i.e., in Deep Convolutional Neural Networks, such as the ones used in (Qawaqneh et al., 2017) toward the useful set of features revealed in this study. Indeed, features humans use are generally more robust for image classification than those used by machine learning algorithms (Geirhos et al., 2018). However, we think that any such further developments will require the inclusion and representation of stimuli and participants from ethnic minority groups, as well as individuals with strong perceptual skills such as super-recognizers. These additions, we believe, will be important steps toward our findings being generalizable, as well as developing high-performing classification algorithms and experimental tools that support forensic and legal decision-making more generally.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72. https://doi.org/10.1038/nature03086

Alink, A., & Charest, I. (2020). Clinically relevant autistic traits predict greater reliance on detail for image recognition. *Scientific Reports*, *10*(1), 14239. https://doi.org/10.1038/s41598-020-70953-8

Bate, S., Bennetts, R., Murray, E., & Portch, E. (2020). Enhanced matching of children's faces in "super-recognisers" but not high-contact controls. *I-Perception*, *11*(4), 2041669520944420. https://doi.org/10.1177/2041669520944420

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. In *Cortex* (Vol. 108, pp. 92–111). https://doi.org/10.1016/j.cortex.2018.07.008

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2018). Facial color is an efficient mechanism to visually transmit emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3581–3586. https://doi.org/10.1073/pnas.1716084115

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://www.ncbi.nlm.nih.gov/pubmed/9176952

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, *17*(14), 7. https://doi.org/10.1167/17.14.7

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43. https://doi.org/10.1016/j.visres.2018.08.004

Edwards, I. (2013). Victims, Sentencing Guidelines, and the Sentencing Council. In *Sentencing*

Guidelines (pp. 71–85). https://doi.org/10.1093/acprof:oso/9780199684571.003.0005

Estéphan, A., Fiset, D., Saumure, C., Plouffe-Demers, M.-P., Zhang, Y., Sun, D., & Blais, C. (2018). Time Course of Cultural Differences in Spatial Frequency Use for Face Identification. *Scientific Reports*, *8*(1), 1816. https://doi.org/10.1038/s41598-018-19971-1

Faghel-Soubeyrand, S., Dupuis-Roy, N., & Gosselin, F. (2019). Inducing the use of right eye enhances face-sex categorization performance. *Journal of Experimental Psychology. General*, *148*(10), 1834–1841. https://doi.org/10.1037/xge0000542

Faghel-Soubeyrand, S., Lecomte, T., Bravo, M. A., Lepage, M., Potvin, S., Abdel-Baki, A., Villeneuve, M., & Gosselin, F. (2020). Abnormal visual representations associated with confusion of perceived facial expression in schizophrenia with social anxiety disorder. *NPJ Schizophrenia*, *6*(1), 28. https://doi.org/10.1038/s41537-020-00116-1

Franqueira, V. N. L., Bryce, J., Al Mutawa, N., & Marrington, A. (2018). Investigation of Indecent Images of Children cases: Challenges and suggestions collected from the trenches. *Digital Investigation*, *24*, 95–105. https://doi.org/10.1016/j.diin.2017.11.002

Gao, F., & Ai, H. (2009). Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method. *Advances in Biometrics*, 132–141. https://doi.org/10.1007/978-3-642-01793-3_14

Geirhos, R., Medina Temme, C. R., Rauber, J., Schütt, H. H., Bethge, M., & Wichmann, F. A. (2018). Generalisation in humans and deep neural networks. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1808.08750

Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *1*, 143. https://doi.org/10.3389/fpsyg.2010.00143

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*, 1–11. https://doi.org/10.1016/j.neuropsychologia.2015.12.004

Goffaux, V., & Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, *6*, 34204. https://doi.org/10.1038/srep34204

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://www.ncbi.nlm.nih.gov/pubmed/11448718

Graw, M., Czarnetzki, A., & Haffner, H. T. (1999). The form of the supraorbital margin as a criterion in identification of sex from the skull: investigations based on modern human skulls. *American Journal of Physical Anthropology*, *108*(1), 91–96. https://doi.org/10.1002/(SICI)1096-8644(199901)108:1<91::AID-AJPA5>3.0.CO;2-X

Home Office. (2017, October 3). *Home Office to crack down on online child sexual abuse with new cutting-edge technology*. GOV.UK. https://www.gov.uk/government/news/home-office-to-crack-down-on-online-child-sexual-abuse-with-new-cutting-edge-technology

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Kau, C. H., & Richmond, S. (2008). Three-dimensional analysis of facial morphology surface changes in untreated children from 12 to 14 years of age. *American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *134*(6), 751–760. https://doi.org/10.1016/j.ajodo.2007.01.037

Kloess, J. A., Woodhams, J., & Hamilton-Giachritsis, C. E. (2021). The challenges of identifying and classifying child sexual exploitation material: Moving towards a more ecologically valid pilot study with digital forensics analysts. In *Child Abuse & Neglect* (Vol. 118, p. 105166). https://doi.org/10.1016/j.chiabu.2021.105166

Kloess, J. A., Woodhams, J., Whittle, H., Grant, T., & Hamilton-Giachritsis, C. E. (2019). The Challenges of Identifying and Classifying Child Sexual Abuse Material. In *Sexual Abuse* (Vol. 31, Issue 2, pp. 173–196). https://doi.org/10.1177/1079063217724768

Komori, M., & Nittono, H. (2013). Influence of Age-independent Facial Traits on Adult

Judgments of Cuteness and Infantility of a Child's Face. *Procedia - Social and Behavioral Sciences*, *97*, 285–291. https://doi.org/10.1016/j.sbspro.2013.10.235

Kramer, R. S. S., Mulgrew, J., & Reynolds, M. G. (2018). Unfamiliar face matching with photographs of infants and children. *PeerJ*, *6*, e5010. https://doi.org/10.7717/peerj.5010

Krasodomski-Jones, A., White, C., Eccleston, D., & Marsh, O. (2017). *Online child sexual abuse imagery*. https://dera.ioe.ac.uk/id/eprint/30926

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Liggett. (1975). The Human Face. By John Liggett. Pp. xxii + 287. (Constable, London, 1974.) Price £3.95. *Journal of Biosocial Science*, *7*(02). https://doi.org/10.1017/s0021932000005848

Merdian, H. L., Perkins, D., McCashin, D., & Stevanovic, J. (2021). Integrating structured individual offending pathway analysis into group treatment for individuals who have accessed, shared, and/or distributed child sexual exploitation material: a feasibility study and preliminary outcome evaluation. In *Psychology, Crime & Law* (Vol. 27, Issue 6, pp. 579–605). https://doi.org/10.1080/1068316x.2020.1849690

Michalski, D., Heyer, R., & Semmler, C. (2019). The performance of practitioners conducting facial comparisons on images of children across age. *PloS One*, *14*(11), e0225298. https://doi.org/10.1371/journal.pone.0225298

Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, *4*, 74. https://doi.org/10.3389/fpsyg.2013.00074

Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1709.01664

Ratnayake, M., Obertová, Z., Dose, M., Gabriel, P., Bröker, H. M., Brauckmann, M., Barkus, A.,

Rizgeliene, R., Tutkuviene, J., Ritz-Timme, S., Marasciuolo, L., Gibelli, D., & Cattaneo, C. (2014). The juvenile face as a suitable age indicator in child pornography cases: a pilot study on the reliability of automated and visual estimation approaches. *International Journal of Legal Medicine*, *128*(5), 803–808. https://doi.org/10.1007/s00414-013-0875-y

Ricanek, K., Bhardwaj, S., & Sodomsky, M. (2015). A review of face recognition against longitudinal child faces. *BIOSIG 2015*. https://dl.gi.de/handle/20.500.12116/2272

Royer, J., Blais, C., Gosselin, F., Duncan, J., & Fiset, D. (2015). When less is more: Impact of face processing ability on recognition of visually degraded faces. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(5), 1179–1183. https://doi.org/10.1037/xhp0000095

Sae-Bae, N., Sun, X., Sencar, H. T., & Memon, N. D. (2014). Towards automatic detection of child pornography. *2014 IEEE International Conference on Image Processing (ICIP)*, 5332–5336. https://doi.org/10.1109/ICIP.2014.7026079

Sforza, C., Grandi, G., De Menezes, M., Tartaglia, G. M., & Ferrario, V. F. (2011). Age- and sex-related changes in the normal human external nose. In *Forensic Science International* (Vol. 204, Issues 1-3, pp. 205.e1–e205.e9). https://doi.org/10.1016/j.forsciint.2010.07.027

Tanner, J. M., & Tanner, J. M. (1990). *Foetus Into Man: Physical Growth from Conception to Maturity*. Harvard University Press. https://play.google.com/store/books/details?id=YxpimctaWd4C

Tardif, J., Fiset, D., Zhang, Y., Estéphan, A., Cai, Q., Luo, C., Sun, D., Gosselin, F., & Blais, C. (2017). Culture shapes spatial frequency tuning for face identification. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(2), 294–306. https://doi.org/10.1037/xhp0000288

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science*, *30*(2), 300–308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/bf03202828

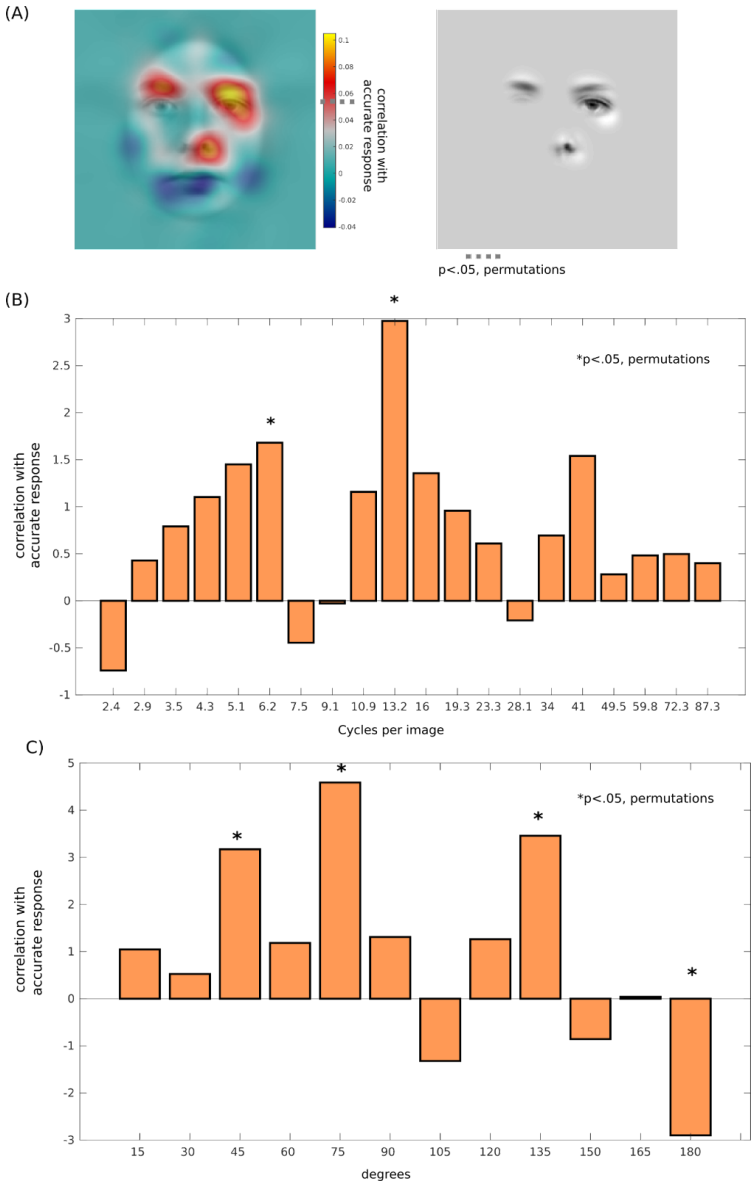
Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Young, S. G., Elliot, A. J., Feltman, R., & Ambady, N. (2013). Red enhances the processing of facial expressions of anger. *Emotion* , *13*(3), 380–384. https://doi.org/10.1037/a0032471

Alliance, W. G. (2020). *Global Threat Assessment 2019: Working together to end the sexual exploitation of children online*. https://apo.org.au/node/278176

Liggett, J. (1974). *The human face*. Stein and Day.

Royer, J., Blais, C., Gosselin, F., Duncan, J., & Fiset, D. (2015). When less is more: Impact of face processing ability on recognition of visually degraded faces. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(5), 1179–1183. https://doi.org/10.1037/xhp0000095



supplementary figure 1. Features for accurate classification of child vs adult faces, irrespective of the class. Correlation between accurate responses and position (A), spatial-frequencies (B) and orientations (C) have been averaged irrespective of the stimulus class. Significance was assessed using permutation testing : a null distribution was created by regressing the presence of each feature type across trials with randomly shuffled the accuracy vector of each participant (repeated 1000 times).

Chapitre 4 – Article 3

One-size-fits-all representation in the best face recognisers

Faghel-Soubeyrand[1,2], S., Alink[3], A., Gosselin, F*[1]. & Charest, I*[1,2].

1. Département de psychologie, Université de Montréal, Canada
2. Centre for Human Brain Health, University of Birmingham, UK
3. University of Hamburg, Faculty of Psychology and Human Movement, Germany

Manuscript in preparation.

Abstract

Faces constitute a great example of realistic, socially relevant, and highly dynamic signal transmission tools. Yet, the current state of knowledge on individual differences in face recognition abilities has surprisingly neglected the fact that the goals of our visual system need to constantly shift to extract different information about faces, e.g. that differentiates identity, gender, or facial-expressions. As a result, we still don't know if good face recognisers perform well across a variety of face processing tasks, and, if so, how they accomplish this feat. Here, we tackled this question by mapping, using reverse correlation, the perceptual representations of faces in two tasks — discrimination of sex and expression from faces — in a cohort of 120 individuals (for a total of 576,000 trials) while concomitantly measuring their face recognition abilities. We found strong correlations between the ability to discriminate sex and facial expressions. Most importantly, we discovered a positive correlation between the ability at the two tasks and the similarity between the perceptual representations of faces used in these tasks. Two factors explain a significant part of these changes, the proximity of an individual's representations to those of the ideal observer, and its proximity to those of the average human representations. These findings provide the first mechanisms explaining how an individual is able to be proficient across face-processing tasks, and suggest that a mixture of both a general, one-size-fits-all representation, and top-down factors contribute to face recognition abilities.

Introduction

The ability to recognise faces is considered to be a stable and heritable trait that varies largely in the general population (Adolphs et al., 2008; Wilmer et al., 2010). The last decade of inquiries have uncovered important individual differences in perceptual (DeGutis et al., 2013; Dunn et al., 2022; Faghel-Soubeyrand et al., 2019; Finzi et al., 2016; Humphreys et al., 2007; Pachai et al., 2013; Tardif et al., 2019) and neural mechanisms (Faghel-Soubeyrand et al., 2022b; Jiahui et al., 2018; McGugin & Gauthier, 2016; Passarotti et al., 2007; Rossion, 2022) supporting this crucial ability. While these studies have pushed forward the idea of general mechanisms behind recognition abilities, such as the use of specific feature information (Pachai et al., 2013; Tardif et al., 2019), differences in face after-effects (Dennett et al., 2012), and variations in similarity with visual and semantic brain computations (Faghel-Soubeyrand et al., 2022a), they have neglected an important constraint imposed on our visual system: our goals constantly shift to match the dynamic world we live in. Thus we must sometimes identify faces, categorise the emotions expressed facially, discriminate the sex of faces, or judge the trustworthiness of our peers, and our brains need to constantly adjust to these different demands by applying different strategies to these tasks (Charbonneau et al., 2020; Dupuis-Roy et al., 2019; Gosselin & Schyns, 2001; Schyns et al., 2020; Smith et al., 2005). This begs the question: are good face recognisers performant across all recognition tasks involving faces? And, if so, how do they accomplish this feat? Even if top-down modulations are a crucial part of cognitive (Kveraga et al., 2007; Wolfe, 1994), neuroscientific (Baldauf & Desimone, 2014; Fan et al., 2020; Hebart et al., 2018; Kay & Yeatman, 2017; Tomita et al., 1999) and machine learning models (Mittal et al., 2020; Richards et al., 2019), the relationships between our ability to modulate complex perceptual representations across tasks and recognition abilities has, to our knowledge, never been explored (with the notable exception of Oruc et al., 2018).

A range of various works, however, can help us make a tentative answer. On the one hand, we could predict from current findings that a one-size-fits-all strategy (e.g. one relying on

the eyes in the horizontals, see Duncan et al., 2019; Pachai et al., 2013; Tardif et al., 2019) is used in skilled recognisers across face recognition tasks. Such general strategies make sense when considering the multiple imperatives dictated by our limited neural apparatus. Indeed, even if the sheer amount of fine-grained visual representations stored in memory can be quite high (Brady et al., 2008; Jenkins et al., 2018), our brain is still inherently limited to a finite set of representations about faces, objects, or any other categories of stimuli we interact with (Barwich, 2019; Quiroga et al., 2005, 2008). Furthermore, a general, flexible visual strategy is plausible considering that representations containing a limited set of features would be sufficient to operate a large number of tasks (Tversky, 1977). Redundancy between visual representations across tasks, then, could emerge from a need for generalisable representations, i.e. for more cost-efficient representations in the human brain (Bonner & Epstein, 2021; Smith et al., 2005). In fact, even recent artificial convolutional neural networks (ANNs) able to accomplish multiple face recognition tasks with top performance (i.e. detect a face, recognise face identities, gender, age, and smiles in Ranjan et al., 2016) can do so using domain-based regularisation, in which the learning parameters are constantly shared across tasks to enable a more general set of features (vs. separately fitting each task-specific parameters). On the other hand, this alternative of similar and shared representations across tasks is difficult to reconcile with the clear differences in information required within each task. Indeed, jumping forward, we will see that the ideal observer — a model observer that uses all possible information optimally to solve a specific task (e.g. Gold et al., 2012; Tjan et al., 1995) — uses very different perceptual representations to discriminate between the sex of our faces (mostly the eye an eyebrow regions; e.g. see Dupuis-Roy et al., 2014) and the emotions that they express (mostly the mouth region; e.g. see Smith et al., 2005). We know that the performance of an observer at a task is necessarily positively correlated with the similarity between his or her perceptual representation and the ideal perceptual representation for this task (Murray, 2011; Murray et al., 2002; Peterson & Eckstein, 2012; Tardif et al., 2019). The pure physical differences in image statistics in a given visual recognition task, thus, dictates that good face recognisers should adapt, at least to a certain degree, their representations to accomplish the

task at hand (Peterson & Eckstein, 2012). This seems to imply that individual performance across face-related tasks will be negatively correlated with the similarity of the perceptual representations in these tasks.

Here, we tackled this question by characterising the relationship between variations in visual content of face representations across two tasks and face recognition abilities in 120 participants (using gold-standard perceptual tests, Russell et al., 2009; see **figure 4.1a**). Specifically, we modelled the mental representations of faces in individuals while they completed two face recognition tasks with identical visual stimulation, discrimination of facial expressions and discrimination of sex. This was done with a data-driven experimental technique similar to Bubbles (Gosselin & Schyns, 2001) that models the categorical representations associated with images based on an efficient subset of visual features (i.e., Gabor Wavelets; see **figure 4.1b** and Alink & Charest, 2020; Faghel-Soubeyrand et al., 2021). The feature space was selected for its similarity with characteristics of receptive fields in the first encoding layers of the visual system (i.e. different spatial frequencies, orientations, spatial coordinates and sizes Daugman, 1980; DeValois et al., 1990; Hubel & Wiesel, 1962; Kay et al., 2008), and account for significant part of the neural code even in higher-order associative regions such as those involved in face and word processing (Kay & Yeatman, 2017). On a given trial, random subsets of this feature space were sampled and presented to a participant. They were asked to categorise these sparse stimuli (e.g. as a male or female in the sex discrimination task), providing a categorical response using computer keys. Each participant saw 2,400 sparse faces per task, for a grand total, for this experiment, of more than 500,000 trials. This sizable dataset enabled us to produce robust quantification, at the single *and* between-participant level, of the association between key visual features we manipulated experimentally and a participants' high-level mental representation of facial expressions and sex (**figure 4.1c**). The visual features characterising such mental representations, which have been the subject of numerous studies in the past 20 years (Adolphs et al., 2005; Dupuis-Roy et al., 2019; Goffaux & Greenwood, 2016; Gosselin & Schyns, 2001, 2002; Pachai et al., 2013; Royer et al., 2018; Schyns et al., 2020; Tardif et al., 2019; Zhan et al., 2019), are

what we refer to as representational content. Understanding how the brain weights specific visual features between different tasks is useful because it gives a clear definition of the information that can be biased by high-level areas when task-demands shift. This also circumvents the difficult step of defining and comparing activity in low/high level brain areas between individuals; surface and spatial coordinates of early V1 area, for example, varies in great scales between individuals (Schwarzkopf et al., 2011). We measured the perceptual representations of each participant in both tasks, separately. By computing the distance in representational contents of participants between these tasks, we obtained a direct measure of similarity between perceptual representations in each of the 120 participants, which we could associate with gold-standard measures for face recognition abilities.

Materials and methods

Participants

A total of 120 healthy volunteers (85 female, mean-age = 21.36, 18-31 years old) with normal or corrected-to-normal vision took part in this experiment. 60 participants were recruited from the University of Birmingham (Birmingham, United Kingdom) and 60 participants were recruited in the greater Montreal area, Canada. Five participants with incomplete data recording and six other participants with abnormal performance (psychophysical thresholds and accuracy 3 SD above the mean of all participants; see *Measures of face-processing abilities* section) were excluded from further analysis. All participants gave their informed consent after being introduced to the experimental procedure in accordance with the Declaration of Helsinki. The experimental procedure was approved by the ethics committee of the University of Birmingham and of the Université de Montréal.

Stimuli : reverse-correlation experiment

The original stimuli used in this study consisted of 16 face images from the Radboud Face dataset (Langner et al., 2010) varying in identity (8 identities), sex (8 male), and facial expressions (8 expressing joy, 8 expressing fear). These faces were converted to 250x250 greyscale images, and were aligned based on twenty landmarks (averaged to six mean coordinates for left and right eyes, left and right eyebrows, nose, and mouth) using Procrustes transformations (rotation, scaling and translation). Each face image was revealed through an ellipsoid mask that excluded non-facial cues. The luminance profile of the resulting face images was equalised across images using the SHINE toolbox (Willenbockel et al., 2010).

Gabors wavelet decomposition

Prior to the experimental phase, we created a feature-space for the 16 face images described in the previous section (i.e. male or female faces expressing either joy or fear by finding a subset of 3000 Gabors wavelets that best-fitted the original image (see Figure 4.1a). The details of the Gabors wavelet decomposition algorithm have been published elsewhere (Alink & Charest, 2020; Faghel-Soubeyrand et al., 2021). Briefly, we used a custom-made Matlab program which aimed to reduce each face image to a subset of the most important 3000 Gabor wavelets (see Figure 4.1a). Gabor wavelets were chosen as our feature-space because they share characteristics with receptive fields in the first encoding layers of the visual system (i.e. different spatial-frequencies, orientations, spatial-coordinates and sizes (Daugman, 1980; DeValois et al., 1990; Hubel & Wiesel, 1962), and account for significant part of the neural code even in higher-order associative regions such as those involved in face and word processing (Kay & Yeatman, 2017). To do so, we considered wavelets of 20 different spatial frequencies exponentially increasing between 2.4 and 76.6 cycles/image (n = [1, 2, …, 20]; $sf = 2 \times 1.2^{n}$), having 18 equidistant orientations

between 0° and 180° (10° to 180° in steps of 10°), and being centred on each possible pixel of the input grayscale image (250^2). The optimally-fitting Gabor features were selected based on their best fit (least-square regression) to the original grayscale image. Amplitudes were set to an equal value for all wavelets. Partial reconstructions of the images at every trial were created by randomly selecting a subset of the 3000 Gabor wavelets and summating them (see **Figure 4.1b**). The pixel intensity range of the resulting images was kept constant, covering the full 0-255 range.

Experimental procedure

We derived the visual content supporting the mental representations of faces in 120 participants across two face-processing tasks. The experimental procedure began after the fitting step described above. Specifically, random subsets of this Gabor Wavelets feature-space were presented to our participants. In each of these trials, participants were asked to categorise the partially sampled image, providing a categorical response using computer keys. The procedure was identical for both tasks; only the experimental instructions (i.e. "categorise the gender of a face", "categorise the emotion of a face) differed. Moreover, the number of gabor features (i.e. the quantity of face information) was adjusted on a trial-by-trial basis during the sex and emotion discrimination tasks to maintain individual performance at an accuracy of 75% correct using the QUEST algorithm (Watson & Pelli, 1983).

Stimuli were presented and behavioural reponses were recorded using Matlab and the Psychtoolbox (Brainard, 1997). Each trial started with a grey screen and a central fixation point (presented for a duration of 250 ms). Afterwards, a partial reconstruction of the facial images (subtending 10° of visual angle) was presented on the central area of the screen and remained on the screen until a button press was made (i.e. either joy vs fear for the emotion discrimination task, or male vs female face for the sex discrimination task). The order of presentation of the 16 original stimuli, as well as the order in which the tasks were completed were randomised. All participants completed 2400 trials of each task. Both tasks were completed over two sessions of

~2 hours each, on separate days. Overall, a total of 576 000 trials were recorded over ~520 hours of testing time.

After the experiment, participants were verbally debriefed and were given a debriefing sheet to take with them. Participants recruited through the Research Participation Scheme of the University of Birmingham were compensated with course credits and those recruited at the University of Montreal and in the greater Montreal area were compensated with course credits or were reimbursed for their time.

Measures of face-processing abilities

Cambridge Face Memory Test long-form (CFMT+). All participants were administered the CFMT long-form, or CFMT+ (Russell et al., 2009). In the CFMT+, participants are required to memorise a series of face identities, and to subsequently identify the newly learned faces among three faces. It includes a total of 102 trials of increasing difficulty. The duration of this test is about 15 minutes. EEG was not recorded while participants completed this test.

Cambridge Face Perception Test. Participants also completed the CFPT (B. Duchaine et al., 2007), which measures the ability to distinguish small differences between simultaneously presented faces. In the CFPT, participants are asked to sort a series of six faces according to their resemblance to a target face. The presented faces are versions of the target face morphed with other faces ranging from 28% and 88% at 6 levels of morphing. The distance between the sequences produced by participants and the correct sequence corresponds to an item's score. Thus, larger scores correspond to greater distances to correct answers and poorer face perception abilities. The CFPT includes trials with inverted faces. Here, only the upright trials were used for our ability scores.

Psychophysical thresholds. The number of Gabor features (i.e. the quantity of face information) was adjusted on a trial-by-trial basis during the face-gender and face-expression tasks as to maintain individual performance at an accuracy of 75% correct using the QUEST algorithm (Watson & Pelli, 1983). Royer et al., (2015; see also Tardif et al., 2019) showed that such a threshold (measured using a similar method during a face identification task) is strongly correlated with three commonly used face recognition ability tests ($r = -0.79$ with the mean of the Cambridge Face Memory Test +, the Cambridge Face Perception Test, and the Glasgow Face Matching Test short version). Accuracies at both tasks were within appropriate range in both tasks ($M_{acc\_emotion} = 75.44\%$, $M_{acc\_gender} = 72.53\%$). We thus used the number of Gabors required to maintain an accuracy of 75% for the categorisation tasks as our individual index of emotion and sex discrimination performance, separately. Average number of Gabor features across participants was 64.83 (2.16% of face information required to attain 75% accuracy) for the emotion discrimination task and 280.37 (9.35% of face information) for the sex discrimination task. Six participants were flagged as outliers because of abnormal accuracies given the procedure (~60%), or response times (< 200 ms or > 10 s).

Data analysis

Dimensionality reduction for feature space

Our psychophysical experiment provided us with a relatively high-dimensional set of features (i.e. 3000 features, per stimulus) to associate with human behavioural judgements about faces. To better extract meaningful differences in representational content from this experimental procedure, we reduced this feature space to a subset of visual features on which all our stimuli could be projected. The psychophysical parameters of this dimensionality reduction algorithm were also chosen to match those of typical *Bubbles* experiments: fiive spatial frequency (SF)

bands (from 0 to 76.6 cycles/image in five logarithmic steps), with varying positions of features within each SF scale (from 2x2 to 8x8 positions in low to high SF bands, in logarithmic steps), and four orientation bands of equal sizes (157.5°-22.5°, 22.5°-67.5°, 67.5°-112.5°, and 112.5°-157.5°).

Feature Diagnosticity Index

This sizable behavioural dataset enabled us to derive robust individual perceptual representations for the sex discrimination task and for the emotion discrimination task at the individual level. We quantified how the same basic image features (see *dimensionality reduction* section) were differently associated with face representations of emotion and gender by computing a Feature diagnosticity index (FDi) for each of the features composing described in the previous section, for every participant and task separately. This amounts, essentially, to computing the dot product between the features matrix presented across trials and the z-scored accuracies. This provided us with two 2D-dimensional matrices of FDi with the dimensions number of participants (109) and number of features (294). Two matrices were thus produced from the same image features that were presented to the participants: one feature matrix weighed by responses for the emotion categorization task, and one weighed by the responses for face-gender categorization task. Example classification images (CI) can be visualised in **Figure 4.1d**, showing the sample-average visual representations for face-sex and face-emotion recognition.

PCA analysis across individual performance scores

A general ability score was obtained by performing a Principal Component Analysis (PCA) on the [120x4] individual scores matrix, and selecting the component projection which explained the most inter-individual variance across ability measures ($PC_{1;}$; see PCA

supplementary figure 4.1). The PC_1 explained 54.7% of face processing ability variations in face-identity memory (CFMT+), face-identity perception (CFPT), sex discrimination performance, and emotion discrimination performance across individuals, and is henceforth referred to as general *f* (see supplementary figure 4.1 for associations between PC_1 and all visual ability tests; $r_{PC1*memorisation}$= 0.6450; $r_{PC1*perception}$= 0.7526; $r_{PC1*emotion}$ = 0.6886; $r_{PC1*sex}$= 0.8555).

Ideal observer models

We also submitted, to both our discrimination tasks, an ideal observer that is, a mathematical model that resolves these tasks optimally. This allowed us to uncover the ideal perceptual representations of these tasks. This was important because it provided us with a benchmark to compare with our observer's representations by defining all physical information available to our observers to categorise specific classes, considering the stimuli of both classes having been shown during the experiment. By the definition of the ideal observer, the closer a human observer's perceptual representation to the ideal observer's for a given task, the better the human observer (Gosselin & Schyns, 2002; Murray, 2005). We computed these ideal observer's perceptual representations by first training linear discriminant models (Barrett *et al.*, 1993) to distinguish the sex or the emotion of our stimuli on a total of 1,280,000 trials for each task (800 trials per stimuli, for 16 stimuli, repeated 100 times), and testing the models on another 1,280,000 trials, deriving cross-validated accuracies at each repetition (5-fold, 5 repetitions, repeated 100 times). When presented with a visual information threshold similar to that of human participants (2.16% of total face information for emotion and 9.35% for gender), these ideal models obtained ceiling accuracies (~99 and 98%, respectively). To force variations in accuracies for these models, we reduced the quantity of visual information presented to 0.5% and 1% for the emotion and the sex discrimination tasks, respectively. These ideal models obtained 84.21% (emotion) and 78.16% (sex) classification accuracy in their respective tasks, and will be referred to as task-ideal observers. Finally, classification Vectors (CVs) for these ideal perceptual representations were derived in an identical manner as with human observers, by weighing the

feature presented across trials of the testing set with the corresponding z-scored accuracies of the model.

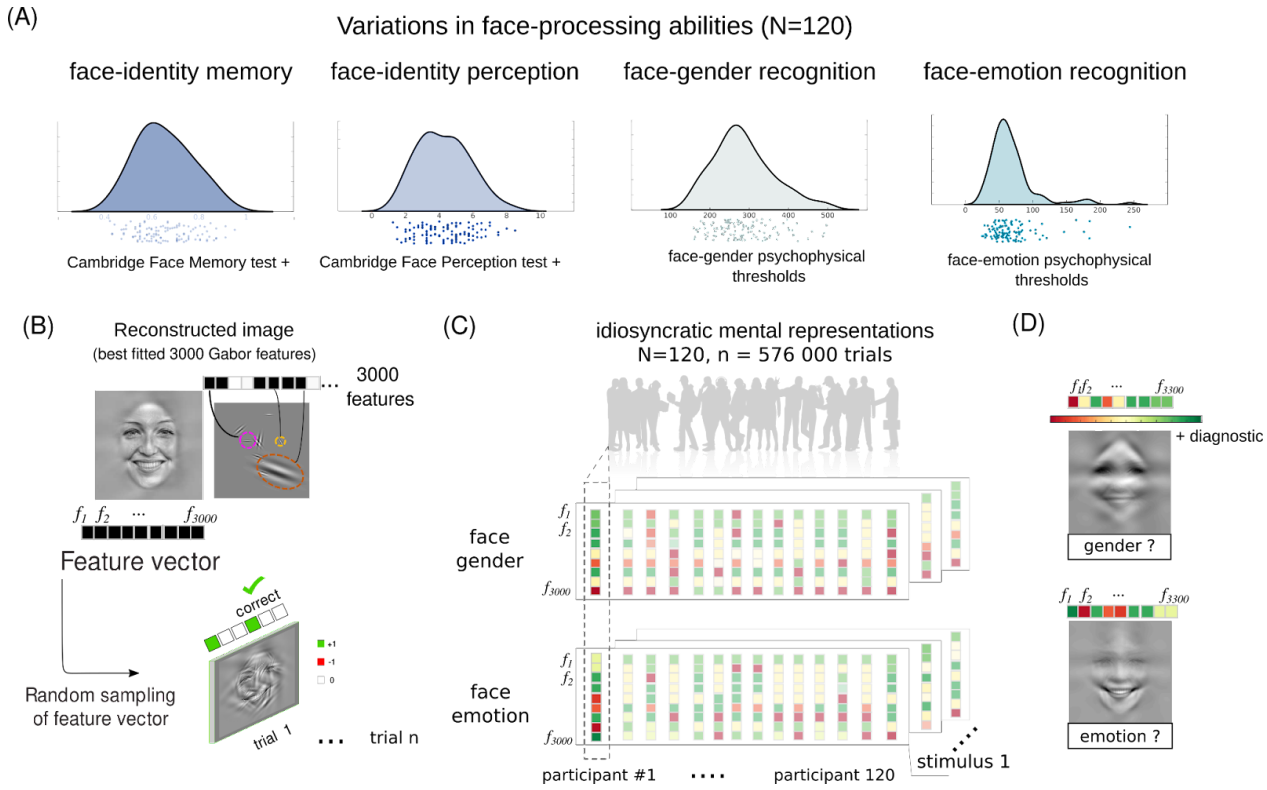


Figure 4.1. Experimental procedures and face-processing ability tests. (A) Variations in face-processing abilities were measured across four tasks using gold-standard psychometric and psychophysical thresholds in 120 participants. (B) To derive visual feature representations in the same 120 participants, we used a reverse-correlation procedure, of which the first step involved the reconstruction of a face image using a subset of 3000 Gabors wavelets varying in spatial-frequency, orientation, x-y coordinates and size. These 3000 wavelets, which best-fitted the original image, are referred to as image features. At any given trial, the stimulus was created by randomly sampling the image features to reveal parts of the reconstructed face image. (C) We derived visual feature representations by asking participants to categorise the resulting sampled faces across many trials in two separate tasks, a face-gender and a face-emotion categorisation task, and associating their behavioural responses to the features shown across trials. Although the stimuli and presentation methods are identical across tasks, this procedure requires that they match the sampled faces with specific mental representations depending on instructions (i.e. joyful or fearful face representations in the emotion task; male or female representations in the gender task). (D) Examples of diagnostic feature representations were obtained for each participant across tasks by weighing the randomly sampled feature vector across trials with standardised behavioural accuracies.

Results

Generalisation of an individual's recognition ability across tasks

We first assessed whether face recognition abilities generalise across tasks. We characterised individual differences in face recognition abilities across tasks by deriving a total of four measures from different face-processing tests. Two perceptual psychophysical thresholds were obtained using the QUEST adaptive staircase procedures (see method), one in a sex discrimination task, the other in an emotion discrimination task. In addition to these thresholds, participants also completed two face recognition tests among the most commonly used in the field: one assessing mnemonic factors in face identitfication (Cambridge Face Memory Test; Duchaine & Nakayama, 2006) and the other assessing perceptual factors in face identification (Cambridge Face Perception Test; Duchaine et al., 2007). Spearman correlation coefficients of individual performance across all tasks are presented on **Supplementary figure 4.2a**. The association between emotion and sex psychophysical thresholds suggests that, indeed, the abiliy in one task generalises, at least partly, to the other (r = .5865, p<.001; **Supplementary figure 4.2b**). Interestingly, we found that an individual' performance at the sex discrimination task also generalises to face identification (r_{CFMT+} = .42 and r_{CFPT} = .49; ps<.001), but that emotion discrimination performance does not [4] (r_{CFMT+} = .14, p =.13; and r_{CFPT} = .34, p<.05; c.f. Oruc et al., 2018). This last finding corroborates some distinction made by models of face recognition indicating similar mechanisms for face identification and sex discrimination, and different mechanisms for face identification and emotion recognition (Calder & Young, 2005; Haxby et al., 2000; Hoffman & Haxby, 2000).

[4] A full factorial multiple linear regression model predicting CFMT+ with age, sex, emotion performance and gender performance as regressors indicated that while face-gender was robustly correlated to face-identification memory and perception (r_{CFMT+} = .52 & r_{CFPT} = .39; p<.001), face-expression recognition was not (ps > .20;c.f. Oruc et al. 2018; total adjusted-r^2 = .1811; $F_{model}(110,4)$ = 7.3524, p<.001).

We further assessed the shared portion of an individual's recognition abilities across face recognition tasks by using Principal Component Analysis (PCA; see Calder et al., 2001; Calder & Young, 2005). The first principal component (PC) of this procedure explained 54.7% of face recognition ability variations across face identification (CFMT+ and CFPT), sex discrimination and emotion dicrimination, and was positively correlated with ability at all these tasks (see PCA supplementary figure for associations between PC_1 and all visual ability tests: $r_{PC1\&CFMT+}= 0.6450$; $r_{PC1\&CFPT}= 0.7526$; $r_{PC1\&emotion} = 0.6886$; $r_{PC1\&sex}= 0.8555$).

Thus, although the lack of correlation between emotion discrimination and face identification abilities partly agrees with the idea that distinct perceptual mechanisms are at the core of these processes, these results suggest that individual face recognition performance generalises across tasks. Indeed, we found that the individual ability to discriminate the sex of faces (Faghel-Soubeyrand et al., 2019) is associated with the ability to discriminate emotions (Duncan et al., 2017, 2019) and with the ability to identify faces (Tardif et al., 2019). Furthermore, we found that a single performance score explains a large part of variance of the performance at each of four face recognition tasks.

There are many possible mechanisms, however, that could explain how an individual's performance in a task might generalise to another. Next, we directly tackle these mechanisms by quantifying an individual's mental representations of faces in a sex discrimination task and an emotion discrimination task.

Revealing the interindividual structure of mental representations across tasks

We computed the representational dissimilarity matrix (RDM) of the perceptual representations in all participants and in the two tasks by computing all pairwise correlational distances between individual Classification-Vectors (CVs, containing 3000 visual features, see **figure 4.1b,c**) derived from our reverse-correlation procedure. This created an interindividual feature RDM (**figure 4.2a**) indicating the geometry across each individual's visual representation within the complete space of our sample and experimental design, i.e. across the same 120

individuals in the face-gender and face-expression tasks. Clear boundaries can be seen between perceptual representations extracted from the two tasks in this RDM. This is also visible in the 2D projection obtained from applying Multidimensional Scaling to this RDM (see **figure 4.2b**) represents the structure of the perceptual representations of our participants. Indeed, perceptual representations of participants within the emotion and the sex discrimination tasks are positively correlated (p<.05, permutations) with averages of ~15% and ~12%, respectively ($min_{emotion}$ = 3.48%, $max_{emotion}$ = 24.35%; $min_{emotion}$ = 6.09%, $max_{emotion}$ = 25.22%), and more so than *between* tasks (5.4% significant positive correlations from face-emotion to face-gender representations; p<.05, permutations; min = 0%, max = 10.6%; paired t-test comparing correlations : $t_{emotion}(112)$ = -16.75,$t_{sex}(112)$ = -12.60 *ps*<.0001). This confirms that feature representations are shared within face-processing tasks and dissociable across face-processing tasks with identical visual stimulation.

Crucially, however, we observed important individual differences in distances between a participant's perceptual representations for the emotion and the sex discrimination tasks (see **figure 4.2C** leftmost panel). As these representational distances can emerge both from meaningful individual differences or noise (Kriegeskorte & Diedrichsen, 2019), we produced new feature representational distances from a feature-space with reduced dimensionality (see method section). With these perceptual representations of reduced dimensionality, we first looked at the relationhip between the correlational distance between a participant's perceptual representations for the two tasks and this participant's overall ability at these tasks. The negative correlation observed (r = -.27, p<.001) indicates that individuals with more similar representations across face processing tasks tend to perform better at face recognition.

We further characterised the relationship between representational distances across tasks and performance by computing the correlational distance between each individual's perceptual representations and the corresponding perceptual representation of the ideal observer (**figure 4.2e**, e.g. $subjectCV_{sex}$ & $idealCV_{sex}$; see method). As we wrote in the introduction, the better a participant the more his or her perceptual representation must be similar to the ideal observer's

(Murray, 2005). This was verified by a negative correlation between the average of these two distances and the average performance at the two face recognition tasks (r = -.25, p<.001; **figure 4.2e**). We also computed the distance between each individual's perceptual representations and the corresponding average perceptual representations of all participants (**figure 4.2d**, i.e. the average task-representation of our sample for a given task). We interpret these group perceptual representations as the physiological pressure on individual perceptual representations. The correlation between the mean of these distances to the group perceptual representations and the average performance at the two tasks is large and negative (r = -.52, p<.001; **figure 4.2d** second panel). A full factorial multiple linear regression model including distances to ideal and group perceptual representations indicates that these two factors are independent, suggesting a role for both in explaining individual differences in face recognition ($F(102,3) = 23.65$, p <.00001; adjusted-$r^2 = 0.3929$; $r_{human\_average} = -0.65$, $p < .00001$; $r_{ideal} = -0.18$, $p = .009$).

Thus, overall, this demonstrates that the more an individual relies on a general, flexible visual representation across tasks that correlates both with the ideal observer and the group perceptual representations, the greater this individual's face recognition ability.

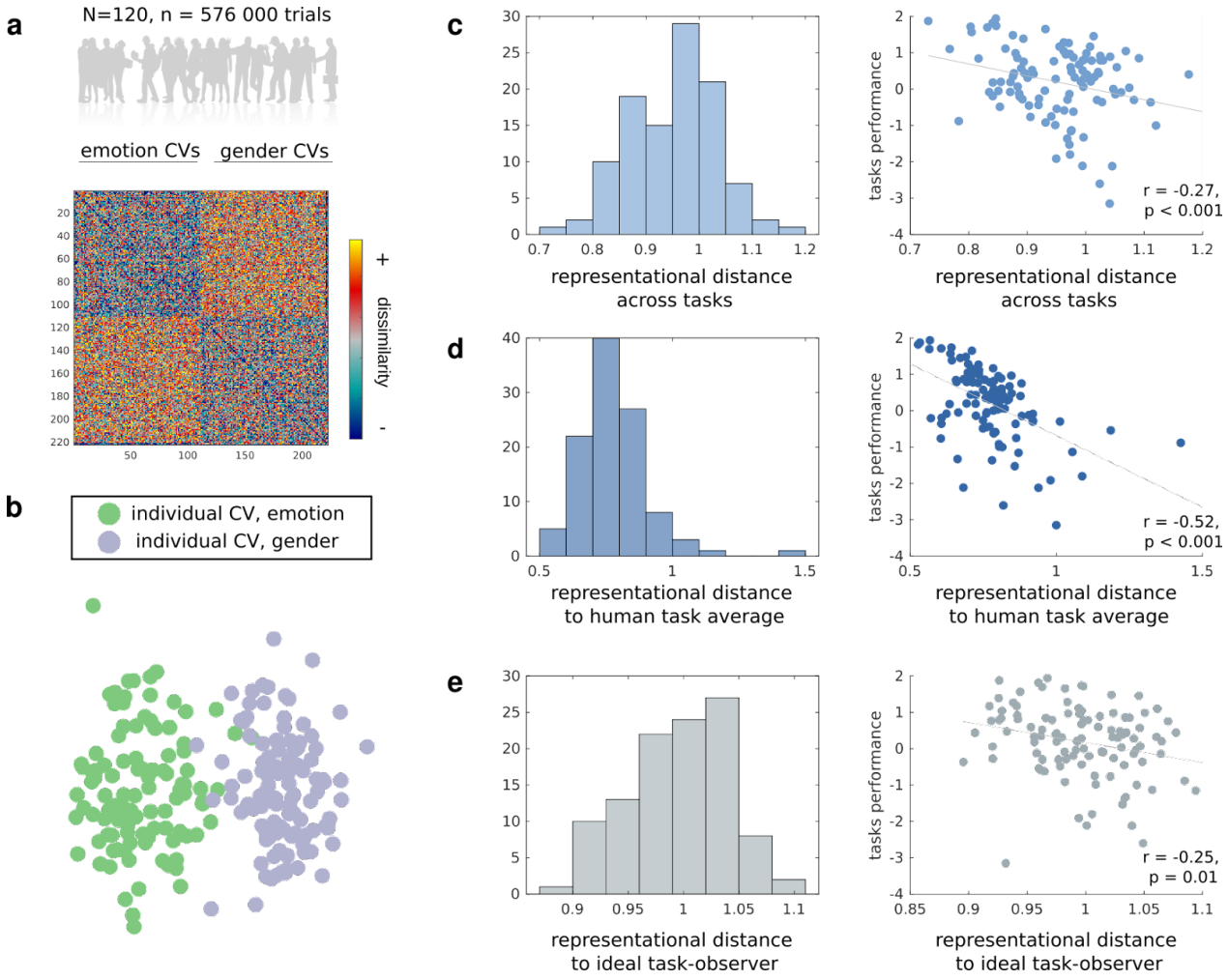


Figure 4.2. Individual differences in visual representations across face-processing tasks associated with individual face-processing abilities. **(a)** Interindividual feature Representational Dissimilarity Matrices (RDMs) were computed by computing the pairwise distance between classification vectors (CVs, see **figure 4.1B**) across participants and tasks, giving a summary of the structure of our participants' feature representations as a function of task-instructions. **(b)** Interindividual feature RDM is projected on a 2D plane using multidimensional scaling. It displays individual CV as single dots for the 109 participants during the face-expression (green dots) and face-gender (purple dots) tasks. **(c)** Representational distances (1- *r*) across tasks were computed for each participant, and correlated with individual recognition performance across the face-expression and face-gender recognition tasks. This indicated that better face recognisers tended to have more similar visual representations across tasks (r = -.27, p<.001). Distances between individual feature representations and **(d)** average human task-representations and **(e)** ideal observer representations are shown with their respective distribution across participants. Individuals with more similar representations to these task representations show better face recognition abilities.

Discussion

Faces constitute a great example of realistic and highly dynamic signal transmission tool with which our visual system interacts daily. Yet, the specific perceptual representations that we use in response to these dynamic demands, as well as how these representations affect our ability to recognise faces, is still largely unknown. Here, we tackled this broad question by recording a large quantity of psychophysical data in 120 individuals during two face recognition tasks —a sex discrimination task and an emotion discrimination task. The specific procedure that we employed enabled us to characterise perceptual representations of faces for each participants *and* for each face recognition task as well as to associate these changes with face recognition abilities. We found strong correlations between the ability to discriminate the sex & emotions, as well as strong correlations between face-sex discrimination and face identification abilities, indicating that individuals can generalise their proficiency across face recognition tasks. Most importantly, we discovered that the more similar a participant's perceptual representations in the two tasks, the better this participant was across these face-processing tasks. This suggests the use of a unique, one-size-fits-all perceptual representation across face recognition tasks in the best face recognisers.

Top-down modulations are crucial for processes ranging from visual search (Wolfe, 1994) to recognition of simple shapes and every-day objects, including faces in both humans (Kay and Yeatman 2017; Fan et al. 2020; Hebart et al. 2018; Kveraga et al. 2007; Baldauf and Desimone 2014) and artificial intelligence models (Mittal et al., 2020; Richards et al., 2019; Yamins & DiCarlo, 2016). Top-down modulations imply the existence of target perceptual representations; where the brain (or algorithm) focuses on relevant visual features (i.e. the targets), however simple or complex, and inhibits task-irrelevant distractors. In humans, this dynamic efficient process is thought to emerge from the modulation of sensory input in posterior sensory brain

areas (e.g. V4 in Moran & Desimone, 1985) by feedback from more anterior neocortical regions (e.g. lateral prefrontal cortex ;Freedman et al., 2003; Harel et al., 2014). Most findings on these phenomena have implicitly assumed that the way our brains complete this process is homogeneous across individuals. This could be due to the use of simple and impoverished stimuli (e.g. sinwave gratings) and well-controlled tasks (e.g. attentional effects between clearly defined targets; (Baldauf & Desimone, 2014), which artificially reduces inter-task and inter-subject variability. The present study characterised individual variations in top-down modulations of realistic representations, i.e. those we rely for face-related information recognition. Our findings indicate significant individual variations in the way we modulate perceptual representations from one task to another, and demonstrate the significance of these differences by associating them to individual face recognition abilities. Doing so, we discovered the first mechanism subtending the proficiency of some individuals across face recognition tasks: a one-size-fits-all strategy that relies on similar visual content across tasks. Top-down modulations of high- to low-level brain areas had been previously linked to higher *within-subject* accuracy (Rutman et al., 2010; Thut et al., 2006), and previous work had been able to show modulation of the actual visual features our brain uses for a specific face representation (e.g. (Gosselin & Schyns, 2001; Jack & Schyns, 2017; Schyns et al., 2020; Smith et al., 2012; Zhan et al., 2019), but none had characterised individual differences in top-down modulations and how they might affect variations in perceptual abilities.

Our findings of a unique perceptual representation in the best face recognisers are reminiscent of (Peterson & Eckstein, 2012), who found that humans, *on average*, tend to first fixate just below the eyes when processing an array of different facial information. Using eye-tracking data and ideal observers with human biases (e.g. foveated information), the authors demonstrated that this unique strategy made the best use of task-relevant information for faces — however without being able to test this claim directly. Indeed, eye tracking does not reveal use of information *per se*, only the *foveated* information of individuals during a task (Arizpe et al., 2012). Here, on the

other hand, our experimental procedure (Alink and Charest 2020; Faghel-Soubeyrand et al. 2021), similar to the Bubbles technique (Gosselin & Schyns, 2001), enabled us to directly extract variations in the perceptual representations of each participant for each face recognition task. Future work will reveal the specific content behind these visual strategies (Dunn et al., 2022).

Why would more variability in perceptual representations between face recognition tasks lead to a diminished face recognition ability? As we wrote in the introduction, we could have easily predicted the opposite. This somewhat surprising finding suggests that structural factors inherent to human physiology and cognition, and not only the physical properties of the visual world, are at play in our ability to recognise face-related information around us. Indeed, if our brains are efficient machines capable of changing goals at will, our memory contains a necessarily limited amount of representations about the outside world (Brady et al., 2008; Jenkins et al., 2018; Quiroga et al., 2008). In this context, changing our representations to perfectly match specific task goals would not be without costs. Instead, it appears that a general, flexible face-representation that is suboptimal in specific tasks, but overall good in many (Dodge & Karam, 2017; Sterkenburg & Grünwald, 2021; Wolpert, 1996), is an efficient compromise between i) our inherently limited biological structure ii) our various evolutionary goals iii) the specific physical/image properties dictated by these goals. Part of this idea is formalised by the no free lunch theorems (NFLT; Wolpert, 1996). In essence, NFLT shows that any optimisation algorithm performing adequately on a given task "A" will perform poorly on another data distribution if there is little overlap between the distributions (Ho & Pepyne, 2002). Thus, it appears that no perfect general-purpose algorithm can exist; to perform, the learning algorithm must either be built in a specialised way to be optimal with the proper problem/task in mind, or built with general, robust parameters that are suited for many purposes.

Conclusion

This work tackled two important questions for the understanding of realistic face cognition that have been neglected from the literature investigating face recognition abilities. The first related to the generalisation of one's ability to process face-related information across different goals: are good face recognisers performant *across* various face processing tasks? We found evidence supporting this. We observed, in a sample of 120 participants, strong correlations between individual ability to recognise sex & their ability to recognise facial expressions, as well as strong correlations between their ability to recognise sex and their ability to identify faces. The second, more arduous question, related to the specific mechanisms behind the ability to be skilled across face-processing tasks : what perceptual strategies do good face-recognisers use to accomplish skilled recognition across tasks? Using a massive reverse correlation dataset collected from the same 120 participants across face-sex and face-expression recognition, we discovered that the best face recognisers rely on *more similar* perceptual content across tasks. We believe that such a mechanism is explained by a compromise between our limited cognitive apparatus, our diverse and ever changing visual goals, and the physical properties required to extract this important information from faces.

Data availability

Data associated with this article will be available online upon publication of the manuscript.

Code availability

The MATLAB codes used in this study will be available online upon publication of the manuscript.

Acknowledgements

Funding for this project was supported by an ERC Starting Grant [ERC-StG-759432] to I.C, by a NSERC grant to F.G., and by NSERC and IVADO graduate scholarships to S.F.S.

Author contributions

(CRediT standardised author statement)

Simon Faghel-Soubeyrand: Conceptualisation, methodology, software, formal analysis, investigation, data curation, writing - original draft, visualisation, project administration, funding acquisition. **A.A.:** Conceptualisation, methodology, software, data curation. **F.G.:** Conceptualisation, methodology, Methodology, formal analysis, writing - original draft, supervision, funding acquisition. **I.C.**: Conceptualisation, methodology, software, supervision, resources, writing - reviewing and editing, funding acquisition.

Competing interests

The authors declare no competing interests.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72. https://doi.org/10.1038/nature03086

Adolphs, R., Spezio, M. L., Parlier, M., & Piven, J. (2008). Distinct face-processing strategies in parents of autistic children. *Current Biology: CB*, *18*(14), 1090–1093. https://doi.org/10.1016/j.cub.2008.06.073

Alink, A., & Charest, I. (2020). Clinically relevant autistic traits predict greater reliance on detail for image recognition. *Scientific Reports*, *10*(1), 14239. https://doi.org/10.1038/s41598-020-70953-8

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: difficulties with eye movements as a measure of information use. *PloS One*, *7*(2), e31106. https://doi.org/10.1371/journal.pone.0031106

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424–427. https://doi.org/10.1126/science.1247003

Baldauf, & Vries. (n.d.). Top-down attention in the face-processing network: an MRI-guided MEG study using multiple simultaneous frequency tags. *Journal of Vision*. https://iris.unitn.it/bitstream/11572/228431/1/VSS_Baldauf_face_freqtag_suppl2_www.pdf

Barwich, A.-S. (2019). The Value of Failure in Science: The Story of Grandmother Cells in Neuroscience. *Frontiers in Neuroscience*, *13*, 1121.

https://doi.org/10.3389/fnins.2019.01121

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature Communications*, *12*(1), 4081. https://doi.org/10.1038/s41467-021-24368-2

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://www.ncbi.nlm.nih.gov/pubmed/9176952

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208. https://doi.org/10.1016/s0042-6989(01)00002-5

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews. Neuroscience*, *6*(8), 641–651. https://doi.org/10.1038/nrn1724

Charbonneau, I., Robinson, K., Blais, C., & Fiset, D. (2020). Implicit race attitudes modulate visual information extraction for trustworthiness judgments. *PloS One*, *15*(9), e0239305. https://doi.org/10.1371/journal.pone.0239305

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, *20*(10), 847–856. https://doi.org/10.1016/0042-6989(80)90065-6

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic

face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, *23*(11), 1279–1287. https://doi.org/10.1177/0956797612446350

DeValois, R. L., Burke, P., De Valois, K. K., & DeValois, K. K. (1990). *Spatial Vision*. OUP USA. https://play.google.com/store/books/details?id=gHPmCwAAQBAJ

Dodge, S., & Karam, L. (2017). A Study and Comparison of Human and Deep Learning Recognition Performance under Visual Distortions. *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, 1–7. https://doi.org/10.1109/ICCCN.2017.8038465

Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: a window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173. https://doi.org/10.1016/j.conb.2006.03.003

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419–430. https://doi.org/10.1080/02643290701380491

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, *17*(14), 7. https://doi.org/10.1167/17.14.7

Duncan, J., Royer, J., Dugas, G., Blais, C., & Fiset, D. (2019). Revisiting the link between

horizontal tuning and face processing ability with independent measures. *Journal of Experimental Psychology. Human Perception and Performance*, *45*(11), 1429–1435. https://doi.org/10.1037/xhp0000684

Dunn, J. D., Varela, V. P. L., Nicholls, V. I., Papinutto, M., White, D., & Miellet, S. (2022). Face-Information Sampling in Super-Recognizers. *Psychological Science*, *33*(9), 1615–1630. https://doi.org/10.1177/09567976221096320

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43. https://doi.org/10.1016/j.visres.2018.08.004

Dupuis-Roy, N., Fiset, D., Dufresne, K., Caplette, L., & Gosselin, F. (2014). Real-world interattribute distances lead to inefficient face gender categorization. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1289–1294. https://doi.org/10.1037/a0037066

Faghel-Soubeyrand, S., Dupuis-Roy, N., & Gosselin, F. (2019). Inducing the use of right eye enhances face-sex categorization performance. *Journal of Experimental Psychology. General*, *148*(10), 1834–1841. https://doi.org/10.1037/xge0000542

Faghel-Soubeyrand, S., Kloess, J. A., Gosselin, F., Charest, I., & Woodhams, J. (n.d.). *Diagnostic Features for Human Categorisation of Adult vs. Child Faces*. https://doi.org/10.31234/osf.io/578bd

Faghel-Soubeyrand, S., Kloess, J. A., Gosselin, F., Charest, I., & Woodhams, J. (2021). Diagnostic Features for Human Categorisation of Adult and Child Faces. *Frontiers in*

Psychology, *12*, 775338. https://doi.org/10.3389/fpsyg.2021.775338

Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022a). The neural code behind real-world recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022b). The neural code behind face recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Fan, X., Wang, F., Shao, H., Zhang, P., & He, S. (2020). The bottom-up and top-down processing of faces in the human occipitotemporal cortex. In *eLife* (Vol. 9). https://doi.org/10.7554/elife.48764

Finzi, R. D., Susilo, T., Barton, J. J. S., & Duchaine, B. C. (2016). The role of holistic face processing in acquired prosopagnosia: evidence from the composite face effect. *Visual Cognition*, *24*(4), 304–320. https://doi.org/10.1080/13506285.2016.1261976

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(12), 5235–5246. https://doi.org/10.1523/JNEUROSCI.23-12-05235.2003

Goffaux, V., & Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, *6*, 34204. https://doi.org/10.1038/srep34204

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The Perception of a Face Is No More Than the Sum of Its Parts. *Psychological Science*, *23*(4), 427–434.

https://doi.org/10.1177/0956797611427407

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://www.ncbi.nlm.nih.gov/pubmed/11448718

Gosselin, F., & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Sciences*, *6*(2), 70–77. https://doi.org/10.1016/s1364-6613(00)01838-6

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(10), E962–E971. https://doi.org/10.1073/pnas.1312567111

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. In *eLife* (Vol. 7). https://doi.org/10.7554/elife.32816

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. https://doi.org/10.1038/71152

Ho, Y.-C., & Pepyne, D. L. (2002). Simple explanation of the no-free-lunch theorem and its implications. *Journal of Optimization Theory and Applications*, *115*(3), 549–570. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.102

3/A:1021251113462&casa_token=RGqLY3vzvUEAAAAA:FEJ5c0SQjqzJCGCv1tHp-bpwGi5KkEisGP8u-A0RGsaRYezj2ziV58bJht379B5PQemtDHRAMgBCNgbvm2A

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *176*(2), 356–373. https://doi.org/10.1007/s00221-006-0621-5

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, *68*, 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings. Biological Sciences / The Royal Society*, *285*(1888). https://doi.org/10.1098/rspb.2018.1319

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), E6418–E6427. https://doi.org/10.1073/pnas.1802246115

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713

Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *eLife*, *6*. https://doi.org/10.7554/eLife.22341

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168. https://doi.org/10.1016/j.bandc.2007.06.007

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

McGugin, R. W., & Gauthier, I. (2016). The reliability of individual differences in face-selective responses in the fusiform gyrus and their relation to face recognition ability. *Brain Imaging and Behavior*, *10*(3), 707–718. https://doi.org/10.1007/s11682-015-9467-4

Mittal, S., Lamb, A., Goyal, A., Voleti, V., Shanahan, M., Lajoie, G., Mozer, M., & Bengio, Y. (2020). Learning to Combine Top-Down and Bottom-Up Signals in Recurrent Neural Networks with Attention over Modules. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2006.16981

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5). https://doi.org/10.1167/11.5.2

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. In *Journal of Vision* (Vol. 2, Issue 1, p. 6). https://doi.org/10.1167/2.1.6

Oruc, I., Shafai, F., & Iarocci, G. (2018). Link Between Facial Identity and Expression Abilities Suggestive of Origins of Face Impairments in Autism: Support for the Social-Motivation Hypothesis. In *Psychological Science* (Vol. 29, Issue 11, pp. 1859–1867). https://doi.org/10.1177/0956797618795471

Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, *4*, 74. https://doi.org/10.3389/fpsyg.2013.00074

Passarotti, A. M., Smith, J., DeLano, M., & Huang, J. (2007). Developmental differences in the neural bases of the face inversion effect show progressive tuning of face-selective regions to the upright orientation. *NeuroImage*, *34*(4), 1708–1722. https://doi.org/10.1016/j.neuroimage.2006.07.045

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Quiroga, R. Q., Quian Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not "Grandmother-cell" coding in the medial temporal lobe. In *Trends in Cognitive Sciences* (Vol. 12, Issue 3, pp. 87–91). https://doi.org/10.1016/j.tics.2007.12.003

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual

representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2016). An All-In-One Convolutional Neural Network for Face Analysis. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1611.00851

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Rossion, B. (2022). Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part II: Neural basis. *Neuropsychologia*, 108279. https://doi.org/10.1016/j.neuropsychologia.2022.108279

Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. In *Cognition* (Vol. 181, pp. 12–20). https://doi.org/10.1016/j.cognition.2018.08.004

Royer, J., Blais, C., Gosselin, F., Duncan, J., & Fiset, D. (2015). When less is more: Impact of face processing ability on recognition of visually degraded faces. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(5), 1179–1183. https://doi.org/10.1037/xhp0000095

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary

face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Rutman, A. M., Clapp, W. C., Chadick, J. Z., & Gazzaley, A. (2010). Early top--down control of visual processing predicts working memory performance. *Journal of Cognitive Neuroscience*, *22*(6), 1224–1234. https://direct.mit.edu/jocn/article-abstract/22/6/1224/4868

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. https://doi.org/10.1038/nn.2706

Schyns, P. G., Zhan, J., Jack, R. E., & Ince, R. A. A. (2020). Revealing the information contents of memory within the stimulus information representation framework. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1799), 20190705. https://doi.org/10.1098/rstb.2019.0705

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring Internal Representations from Behavioral and Brain Data. In *Current Biology* (Vol. 22, Issue 3, pp. 191–196). https://doi.org/10.1016/j.cub.2011.11.061

Sterkenburg, T. F., & Grünwald, P. D. (2021). The no-free-lunch theorems of supervised learning. *Synthese*, *199*(3), 9979–10015. https://doi.org/10.1007/s11229-021-03233-1

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., &

Gosselin, F. (2019). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science*, *30*(2), 300–308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(37), 9494–9502. https://doi.org/10.1523/JNEUROSCI.0875-06.2006

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. In *Vision Research* (Vol. 35, Issue 21, pp. 3053–3069). https://doi.org/10.1016/0042-6989(95)00070-g

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*(6754), 699–703. https://doi.org/10.1038/44372

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/bf03202828

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11),

5238–5241. https://doi.org/10.1073/pnas.0913053107

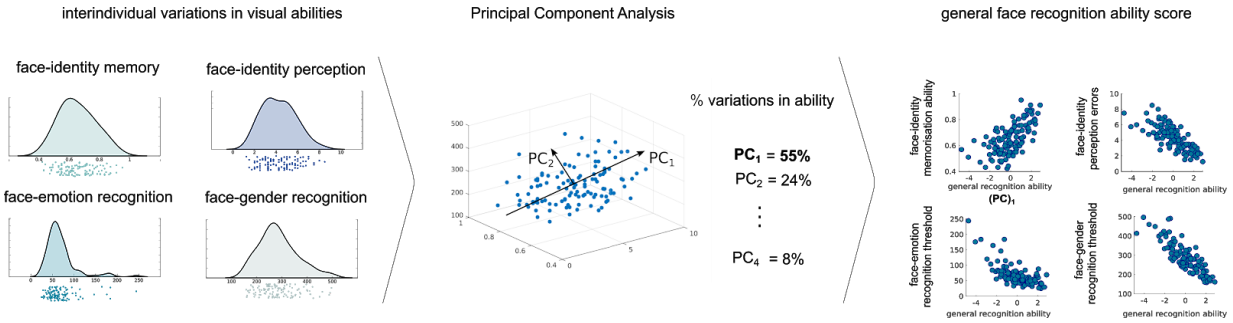
Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1341

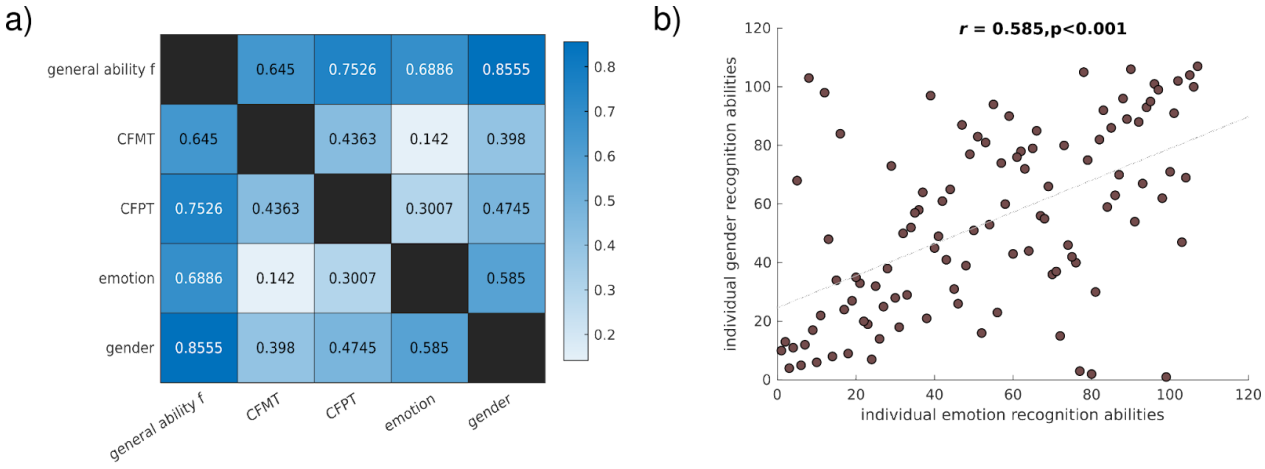
Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

Zhan, J., Garrod, O. G. B., van Rijsbergen, N., & Schyns, P. G. (2019). Modelling face memory reveals task-generalizable representations. *Nature Human Behaviour*, *3*(8), 817–826. https://doi.org/10.1038/s41562-019-0625-3

Supplementary material



Supplementary figure 4.1. Measuring variations in face-recognition abilities and extracting a general face recognition ability component with principal component analysis.



Supplementary figure 4.2. Associations between abilities across face processing tasks.

Chapitre 5 – Article 4

Decoding face recognition abilities in the human brain

Simon Faghel-Soubeyrand[1,2], Meike Ramon[3], Eva Bamps [5], Matteo Zoia [6], Jessica Woodhams [2], Anne-Raphaelle Richoz[4], Roberto Caldara[4], Frédéric Gosselin*[1] & Ian Charest*[1,2]

1. Département de psychologie, Université de Montréal, Canada
2. School of Psychology, University of Birmingham, UK
3. Applied Face Cognition Lab, Institute of Psychology, University of Lausanne, Switzerland
4. Department of Psychology, University de Fribourg, Switzerland
5. Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Belgium
6. Department for Biomedical Research, University of Bern, Switzerland

Manuscript currently under review at *Nature Communications*. **Preprint :** Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022). The neural code behind face recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Author contributions
(CRediT standardised author statement)

Simon Faghel-Soubeyrand: Conceptualisation, methodology, software, formal analysis, investigation, data curation, writing - original draft, visualisation, supervision, project administration, funding acquisition.

M.R.: Investigation, resources, project administration, writing - review and editing. **E.B.:** investigation, project administration. **M.Z.**: investigation. **J.W.:** funding acquisition, writing - review and editing. **A-R.R.**: Investigation. **R.C.:** Resources. **F.G.:** Methodology, writing - original draft, supervision, funding acquisition. **I.C.**: Supervision, methodology, software, resources, formal analysis, writing - original draft, project administration, funding acquisition.

Word counts. Abstract: 153. Main text : 2204, Method : 2214

Abstract

Understanding the neural mechanisms supporting face recognition ability has proven elusive. To tackle this challenge, we used a multi-modal data-driven approach combining neuroimaging and behavioural tests with state-of-the-art computational modelling. We recorded the high-density electroencephalographic brain activity of super-recognisers and typical recognisers in response to visual stimuli. Using multivariate pattern analyses, we decoded face recognition abilities from 1 second of brain signals with up to 80% sensitivity. To better understand the mechanisms subtending this decoding, we compared brain computations with computations in artificial models of vision and semantics. We observed larger correlations in the brains of super-recognisers with mid-level vision and with semantic computations. Semantic computations are a component of prominent face recognition models, and this provides the first empirical evidence for their associations with face recognition abilities. Using such multi-modal data-driven approaches is likely to play a critical role in further revealing the complex nature of idiosyncratic face recognition in the human brain.

Introduction

The ability to robustly recognise the faces of our colleagues, friends and family members is paramount to our success as social beings. Our brains complete this feat with apparent ease and speed, in a series of computations unfolding within tens of milliseconds in a wide brain network comprising the inferior occipital gyrus, the fusiform gyrus, the superior temporal sulcus, and more anterior areas such as the anterior temporal lobe (Duchaine & Yovel, 2015; Grill-Spector et al., 2017; Kanwisher et al., 1997). Accumulating neuropsychological and behavioural evidence indicates that not all individuals, however, are equally competent at recognising faces in their surroundings (White and Burton 2022). Developmental prosopagnosics show a great difficulty at this task despite an absence of brain injury (Susilo & Duchaine, 2013). In contrast, super-recognisers exhibit remarkable abilities for processing facial identity, and can recognize individuals even after little exposure several years before (Noyes et al., 2017; Ramon, 2021; Russell et al., 2009). The specific nature of the neural processes responsible for these individual differences remains largely unknown.

So far, individual differences studies have used univariate techniques to investigate face-specific aspects of brain processing. This revealed that contrasts between responses to faces compared to non-faces, measured by the N170 event-related potential component or by the blood oxygen level dependent signals in regions of interest, are modulated by ability (Elbich and Scherf 2017; Herzmann et al. 2010; Huang et al. 2014; Kaltwasser et al. 2014; Lohse et al. 2016; Rossion et al. 2020; Nowparast Rostami et al. 2017). However, univariate and contrast approaches are limited in their capacity to reveal the precise nature of the underlying brain computations (Vinken et al. 2022; Visconti di Oleggio Castello et al. 2021; Dwivedi et al. 2021; Harel et al. 2013).

Here, we tackled this challenge with a data-driven approach. We examined the functional differences between the brains of super-recognisers and typical recognisers using decoding and representational similarity analyses (RSA; Kriegeskorte et al. 2008; Kriegeskorte and Kievit

2013; Charest et al. 2014; Dwivedi et al. 2021; Kriegeskorte and Diedrichsen 2019) applied to high-density electrophysiological (EEG) signals and artificial neural network models. We recruited 33 participants, including 16 super-recognisers, i.e., individuals better than the 98th percentile on a battery of face recognition tests (Russell et al., 2009; **Fig. 5.1a**). We measured EEG in more than 100,000 trials while participants performed a one-back task. The objects depicted in the stimuli belonged to multiple visual categories including face images of different sexes, emotions, and identities, as well as images of man-made and non-face natural objects (e.g., a computer, a plant), animals (e.g., a giraffe, a monkey), and scenes (e.g., a city, a dining room; **Fig. 5.1b**).

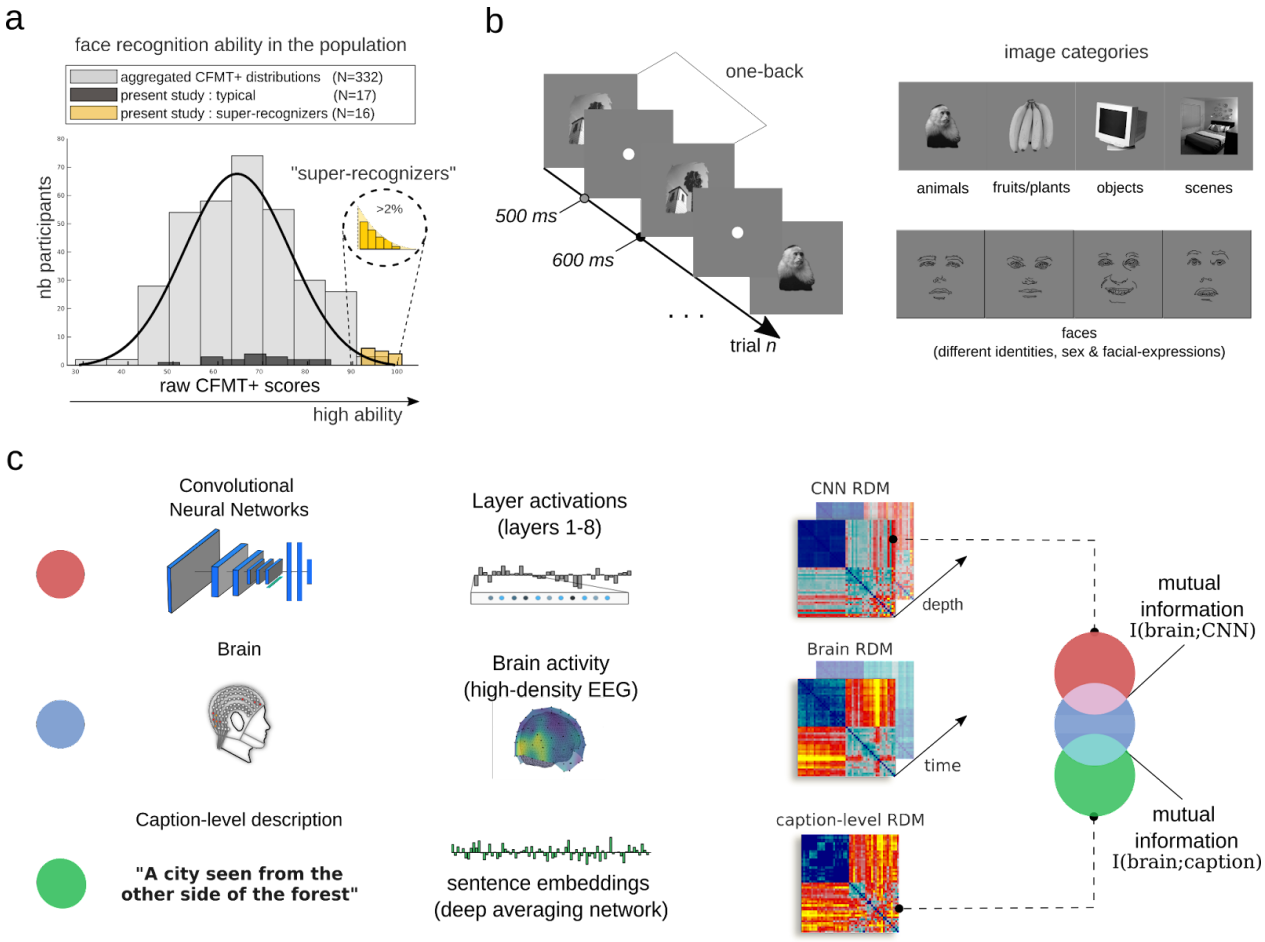


Figure 5.1. Experimental procedure and computational modelling of brain representations. a) The histogram shows the Cambridge Face Memory Test long-form (CFMT+, Russell et al. 2009) scores of

super-recognisers (yellow bars), typical recognisers (black bars), and an additional 332 neurotypical observers from three independent studies for comparison (Faghel-Soubeyrand et al., 2019; Fysh et al., 2020; Tardif et al., 2019). **b)** Participants engaged in a one-back task while their brain activity was recorded with high-density electroencephalography. The objects depicted in the stimuli belonged to various categories, such as faces, objects, and scenes. Note that the face drawings shown here are an anonymised substitute to the experimental face stimuli presented to our participants. **c)** Representational dissimilarity matrices (RDM) were computed from convolutional neural networks (CNN; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014) of vision, human brain activity, and a deep neural network of caption classification and sentence semantics (Cer et al., 2018). To characterise the CNN RDMs, we computed the pairwise similarity between unit activation patterns for each image independently in each CNN layer. The caption-level RDMs were derived from human caption descriptions of the images transformed into sentence embeddings. Brain RDMs were computed using cross-validated decoding performance between the EEG topographies from each pair of stimuli at every 4 ms time-point. Mutual information (Ince et al., 2017) between the model RDMs and the brain RDMs was assessed, for every participant, at each 4 ms step from stimulus onset.

Results

Discriminating super-recognisers and typical recognisers from 1 second of brain activity

With this sizable and category-rich dataset, we first attempted to classify a participant as either a super- or a typical recogniser based solely on their brain activity. More specifically, we trained Fisher linear discriminants to predict group membership from single, 1-second trials of EEG patterns (in a moving searchlight of five neighbouring electrodes). We observed up to ~80% cross-validated decoding performance, peaking over electrodes in the right hemisphere. This performance is impressive given that the noise ceiling imposed on our classification by the test-retest reliability of the CFMT+ (Duchaine and Nakayama 2006; Russell et al. 2009), the gold-standard test used to identify super-recogniser individuals, is ~93% (SD=2.28%; see Supplementary material). To reveal the time course of these functional differences, we applied the same decoding procedure to each 4-ms interval of EEG recordings. Group-membership predictions attained statistical significance (p<.001, permutation tests, **Fig. 5.2a**) from about 65 ms to at least 1 s after stimulus onset, peaking around 135 ms, within the N170 window (Bentin et al., 1996; Rossion & Jacques, 2012). Notably, similar results were obtained following the presentation of both face *and* non-face visual stimuli (**Fig. 5.2a; see also Supplementary Fig.**

5.1). The decoding of group-membership from non-face stimuli could be due to face features stored in short-term memory from one-back trials. To control for this possibility, we repeated our decoding analysis for non-face trials either preceded by face trials or by non-face trials. We found significant decoding of group membership in both cases (**Supplementary figure 5.1**). Altogether, these results corroborate a central prediction of domain-general accounts of face recognition (Behrmann & Avidan, 2005; Garrido et al., 2018; Geskin & Behrmann, 2018; Grill-Spector et al., 2004; Kanwisher, 2000).

Predicting recognition ability from 1 second of brain activity

An ongoing debate in individual differences research is whether the observed effects emerge from qualitative or quantitative changes in the supporting brain mechanisms (Barton and Corrow 2016; Bobak et al. 2017; Rosenthal et al. 2017; Hendel et al. 2019; Vogel et al. 2005; Maguire et al. 2003; Zadelaar et al. 2019; Price and Friston 2002; Anderson et al. 2020). The decoding results presented up to this point might give the impression that face recognition ability is supported by qualitative differences in brain mechanisms. However, these results were obtained with dichotomous classification models applied, by design, to the brains of individuals from a bimodal distribution of ability scores (e.g. Maguire et al., 2003).

To better assess the nature of the relationship between neural representations and ability in the general population, we thus performed a new decoding analysis on the typical recognisers only, using a continuous regression model. Specifically, we used cross-validated fractional ridge regression (Rokem & Kay, 2020) to predict *individual* CFMT+ face recognition ability scores from single-trial EEG data. This showed essentially similar results to the previous dichotomic decoding results: performance was above statistical threshold (p<.01, FDR-corrected) from about 80 ms to at least 1 s, peaking around 135 ms following stimulus onset for both face and non-face stimuli (**Fig. 5.2b,** peak-rho$_{face}$=.4149 at 133 ms, peak-rho$_{non\text{-}face}$=.4899 at 141 ms). This accurate decoding of individual scores from EEG patterns is compatible with a quantitative account of variations in brain mechanisms across individuals differing in face recognition abilities.

Altogether, these decoding results provide evidence for important, domain-general, quantitative and temporally extended variations in the brain activity supporting face recognition abilities. This extended decoding suggests effects of individual ability across multiple successive processing stages.

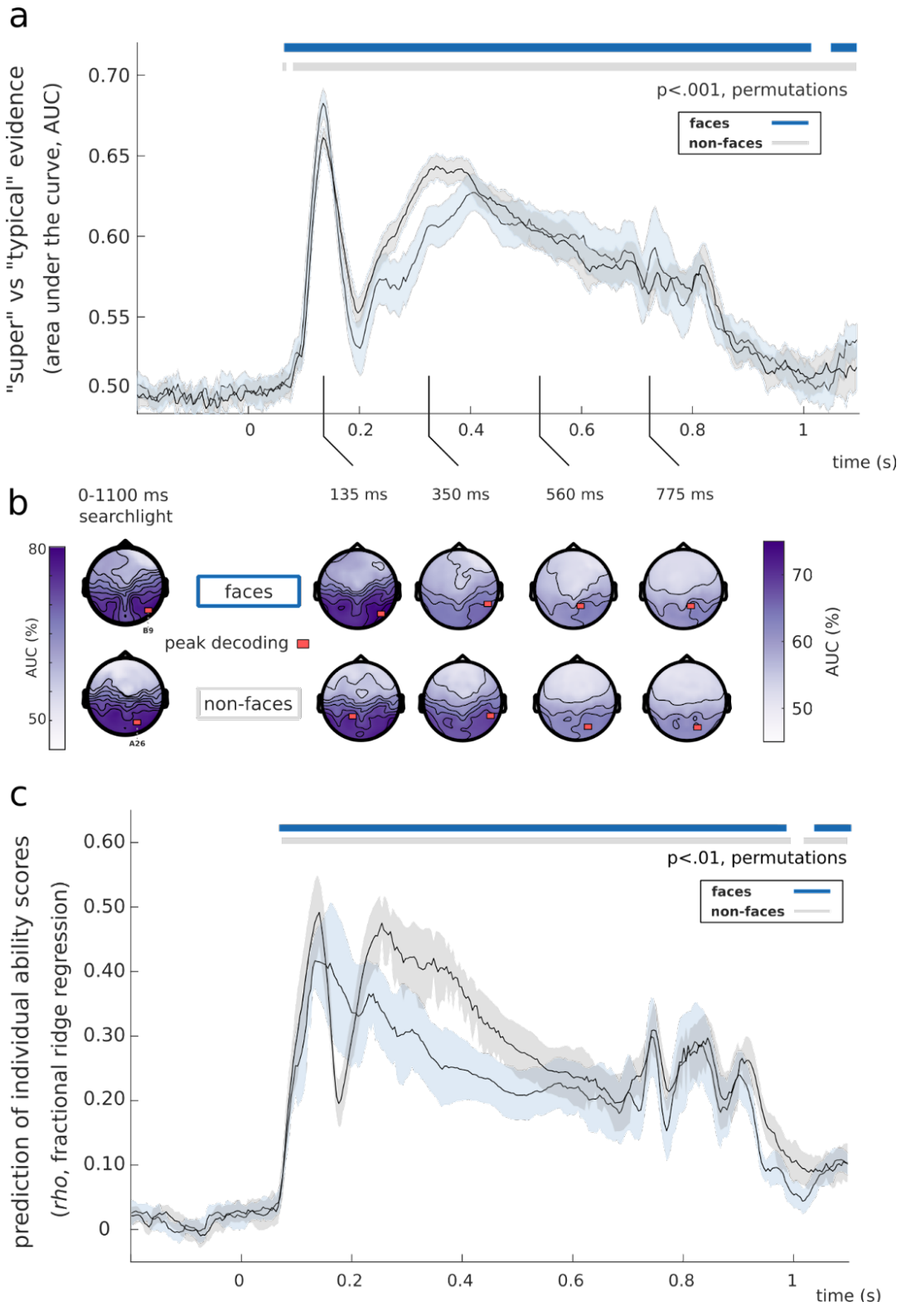

a
"super" vs "typical" evidence
(area under the curve, AUC)
0.70
0.65
0.60
0.55
0.50
p<.001, permutations
faces
non-faces
0
0.2
0.4
0.6
0.8
1
time (s)
b
0-1100 ms
searchlight
135 ms
350 ms
560 ms
775 ms
80
AUC (%)
50
faces
peak decoding
non-faces
B9
A26
70
60
50
AUC (%)
c
prediction of individual ability scores
(rho, fractional ridge regression)
0.60
0.50
0.40
0.30
0.20
0.10
0
p<.01, permutations
faces
non-faces
0
0.2
0.4
0.6
0.8
1
time (s)

Figure 5.2. Decoding interindividual recognition ability variations from EEG activity. a) Trial-by-trial group-membership predictions (super-recogniser or typical recogniser) were computed from EEG patterns, for each 4-ms interval, while participants processed face (blue trace) or non-face stimuli (grey trace). Significant decoding performance occurred as early as 65 ms, peaked in the N170 window, and lasted for the remainder of the EEG epochs (p<.001). **b)** Topographies were obtained using searchlight decoding analyses, either concatenating all time points (left topographies) or for selected time-windows (right topographies). Concatenating all time points resulted in peak classification performance of 77.3% over right occipito-temporal electrodes for face and 77.5% over right occipito-temporal electrodes for non-face conditions. In the N170 window, we observed a peak classification performance of 74.8% over right-temporal electrodes for face, and 72.1% over left-temporal electrodes for non-face conditions. **c**) We decoded the CFMT+ scores of the typical recognisers using fractional ridge regression. This yielded similar results with significant decoding as early as 75 ms, peaking around the N170 time window (peak-rho_{face}=.4149, peak-$rho_{non\text{-}face}$=.4899), and lasted for the remainder of the EEG epochs (p<.01, 1K permutations, 10 repetitions).

Linking neural representations to computational models of vision

Decoding time courses, and evidence for domain generality, however, offer limited insights on the level of brain computations (Lamme & Roelfsema, 2000; McDermott et al., 2002). To better characterise the visual brain computations covarying with face recognition ability, we compared, using representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008; Kriegeskorte & Kievit, 2013; Charest et al., 2014), the brain representations of our participants to that of convolutional neural networks (CNNs) trained to categorise objects (Krizhevsky et al. 2012; Simonyan & Zisserman, 2014; Güçlü & van Gerven, 2015). These CNNs process visual features of gradually higher complexity and abstraction along their layers (Güçlü & van Gerven, 2015), from low-level (e.g., orientation, edges) to high-level features (e.g., objects and object parts).

The brain representations were characterised by computing representational dissimilarity matrices (RDMs) for each participant and for each 4-ms time interval. These brain RDMs were derived using the cross-validated decoding performance of a linear discriminant model, where brain activity was decoded for every pair of stimuli at a given time interval (Carlson et al., 2013; Cichy et al., 2014; see **Supplementary Fig. 5.2** for the group-average RDMs and time course of key categorical distinctions). The visual model representations were characterised by computing

RDMs from the layers of the CNNs, using Pearson correlations of the unit activations across all pairs of stimuli. Compared to typical participants, we found that the brain RDMs of super-recognisers showed larger mutual information (Ince et al., 2017) with the layer RDMs of CNNs that represent mid-level features (e.g., combinations of edges, contour, shape, texture; Güçlü & van Gerven, 2015; Long et al., 2018) between 133 and 165 ms (**Fig. 5.3a**, p<.05, cluster-test; see also **Supplementary Fig. 5.3**).

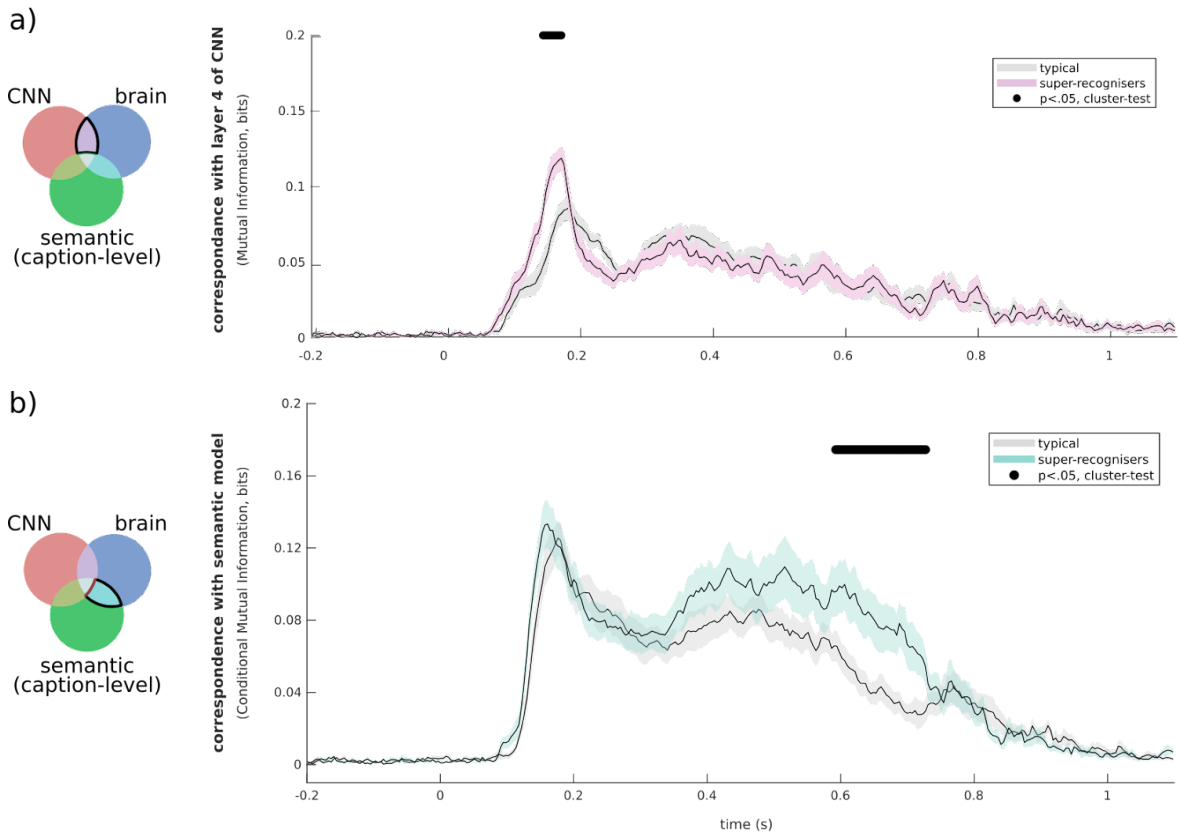


Figure 5.3. Comparison of super- and typical-recogniser brain representations with those of artificial neural networks of visual and semantic processing. a) Mutual information between brain RDMs and AlexNet RDMs (removing shared mutual information between brain and semantic model) is shown for typical- (grey curve) and super-recognisers (pink curve). We found greater similarity with mid-level visual computations (layer 4 shown, but similar results for layers 3 and 6 and for mid-layers of VGG16, another popular CNN model; see **Supplementary Fig. 5.3**) in the brains of super-recognisers (black line indicates significant contrasts, p<.05, cluster-corrected) between 133 ms and 165 ms. Similar results were observed when comparing brains and CNN models without removing the shared mutual information between brains and the semantic (caption-level) model (**Supplementary Fig. 5.3**). **b**) Mutual Information with the semantic model (excluding shared mutual information between brain and AlexNet)

differed for typical- and super-recognisers in a later time window centred around 650 ms (cyan curve; super > typical, $p < .05$, cluster-corrected). Again, similar results were observed when comparing brains and the semantic model without removing the shared mutual information between the brain and AlexNet (**Supplementary Fig. 5.3**). The shaded areas of all curves represent the standard error.

Linking neural representations to a computational model of semantics

The finding that ability decoding was significant as late as 1 s after stimulus onset hints that brain computations beyond what is typically construed as pure visual processing also differ as a function of face recognition ability. To test this hypothesis, we asked five new participants to write captions describing the images presented during our experiment (e.g., "A city seen through a forest."), and used a deep averaging network (Google Universal Sentence Encoder, GUSE; Cer et al., 2018) to transform these captions into embeddings (points in a caption space). GUSE has been trained to predict semantic textual similarity from human judgments, and its embeddings generalise to an array of other semantic judgement tasks (Cer et al., 2018). We then compared the RDMs computed from this semantic model to the brain RDMs of both typical- and super-recognisers. Importantly, both this comparison, and the one comparing brain and visual models, excluded the information shared between the semantic and visual models (but see **Supplementary Fig. 5.3** for similar results with unconstrained analyses). We found larger mutual information with these semantic computations in the brains of super-recognisers than in those of typical recognisers in a late window between 598 and 727 ms (**Fig. 5.3b**, $p<.05$, cluster-test).

Discussion

We investigated the neural mechanisms behind face recognition ability variations using a data-driven approach combining neuroimaging, behavioural tests, and computational models. We recorded the high-density electroencephalographic (EEG) response to face and non-face stimuli in super-recognisers and typical recognisers. We reliably decoded group membership as well as

recognition abilities of single individuals from 1 second of brain activity. To characterise the neural computations underlying these individual differences, we compared human brain computations with those of artificial neural network models of vision and semantics. This revealed stronger associations between early brain representations of super-recognisers and mid-level computations of vision models, and stronger associations between late brain representations of super-recognisers and semantic model computations. To our knowledge, this is the first demonstration of a link between face recognition ability and brain computations beyond high-level vision.

We achieved robust decoding of face recognition ability when examining EEG responses to face *and* non-face stimuli. This domain-general decoding result indicates that mechanisms underlying face recognition abilities give rise to enhanced neural representations that are not restricted to faces (Vinken et al. 2022). This is consistent with several neuropsychological (Barton et al., 2019; Bobak et al., 2016; Duchaine et al., 2007; Gabay et al., 2017; Geskin & Behrmann, 2018; Hendel et al., 2019) and brain imaging findings (Avidan et al., 2005; Jiahui et al., 2018; Kaltwasser et al., 2014; Rosenthal et al., 2017) showing face and non-face processing effects in individuals across the spectrum of face recognition ability (Behrmann & Plaut, 2013; Harel et al., 2013; but see Duchaine et al., 2006; Furl et al., 2011; Lohse et al., 2016; Wilmer et al., 2012). In addition, this decoding approach may provide quick and accurate alternatives to standardised behavioural tests assessing face recognition ability, for example in the context of security settings that benefit from strong face processing skills among their personnel (such as police agencies, border patrol, etc.). It could also be used in a closed-loop training procedure designed to improve face recognition ability (Faghel-Soubeyrand et al. 2019).

The decoding we observed for face and non-face stimuli peaked in the temporal window around the N170 component (Bentin et al., 1996). At that time, the computations in the brains of our participants differed most with respect to the mid-layer computations of artificial models of vision. These layers have been previously linked to processing in human infero-temporal cortex (hIT; Khaligh-Razavi & Kriegeskort,e 2014; Güçlü & van Gerven, 2015; Jiahui et al. 2018) and

functionally to mid-level feature representations such as combinations of edges and parts of objects (Güçlü & van Gerven, 2015; Long et al., 2018). This does not mean that the N170 is exclusively involved in these mid-level processes. Rather, it suggests that other visual computations, including the high-level visual computations usually associated with the N170, do not differ substantially between super-recognisers and typical recognisers. The fact that these mid-level features are mostly shared between face and non-face stimuli could explain at least partly the high decoding performance observed for both classes of stimuli.

Finally, we found that face recognition ability is also associated with semantic computations that extend beyond basic-level visual categorisation in a late time-window around the P600 component (Eimer et al., 2012; Shen et al., 2016; van Herten et al., 2005). Recent studies using computational techniques have shown that word representations derived from models of natural language processing explain significant variance in the visual ventral stream (Dwivedi et al., 2021; Fernandino et al., 2022; Popham et al., 2021). The current study goes beyond this recent work in two ways. First, our use of human sentence description and sentence encoders to characterise semantic (caption-level) computations provides a more abstract description of brain representations. Second, and most importantly, our work revealed a link between semantic brain computations and individual differences in face recognition ability. An association between semantic processes and face recognition ability had been posited in models of face recognition (Bruce & Young, 1986; Duchaine & Yovel, 2015) but, to our knowledge, it had never been shown empirically before.

With the development of novel and better artificial models simulating an increasing variety of cognitive processes, and with the technological advances allowing the processing of increasingly larger neuroimaging datasets, the approach described here provides a stepping stone for better understanding face recognition idiosyncrasies in the human brain.

Methods

Participants

A total of 33 participants were recruited for this study. The first group consisted of 16 individuals with exceptional ability in face recognition — super-recognisers. The second group was composed of 17 neurotypical controls. These sample sizes were chosen according to the effect sizes described in previous multivariate object recognition studies (Carlson et al., 2013; Cichy et al., 2014; Hebart et al., 2018). The data from one super-recogniser was excluded due to faulty EEG recordings. No participant had a history of psychiatric diagnostic or neurological disorder. All had normal or corrected to normal vision. This study was approved by the Ethics and Research Committee of the University of Birmingham, and informed consent was obtained from all participants.

Sixteen previously known super-recognisers were tested in the current study (30-44 years old, 10 female). Eight of these (SR1-SR8) were identified by Prof. Josh Davis from the University of Greenwich using an online test battery comprising a total of six face cognition tasks (Noyes et al., 2021) and tested at the University of Birmingham. The remaining eight (SR-9 to SR-16) were identified using three challenging face cognition tests (Ramon, 2021) and were tested at the University of Fribourg. The behavioural test scores for all participants are provided in **Supplementary Tables 1 and 2.** Across SR cohorts, the Cambridge Face Memory Test long-form (CFMT+; Russell et al., 2009) was used as the measure of face identity processing ability. A score greater than 90 (i.e., 2 SD above average) is typically considered the threshold for super-recognition (Bobak et al., 2016; Davis et al., 2016; Russell et al., 2009). Our 16 super-recognisers all scored above 92 (M=95.31; SD=2.68). A score of 92 corresponds to the 99th percentile according to our estimation from a group of 332 participants from the general population recruited in three independent studies (Faghel-Soubeyrand et al., 2019; Fysh et al.,

2020; Tardif et al., 2019).

An additional 17 typical recognisers (20-37 years old, 11 female) were recruited and tested on campus at the University of Fribourg (n=10) and the University of Birmingham (n=7). Their CFMT+ scores ranged from 50 to 85 (M=70.00; SD=9.08). Neither the average nor the distribution of this sample differed significantly from those of the 332 participants from the general population mentioned above (see **Fig. 5.1a**; t(346)=1.3065, p=0.1922; two-sample Kolmogorov-Smirnov test; D(346)=0.2545, p=0.2372).

Tasks

CFMT+

All participants were administered the CFMT long-form, or CFMT+ (Russell et al., 2009). In the CFMT+, participants are required to memorise a series of face identities, and to subsequently identify the newly learned faces among three faces. It includes a total of 102 trials of increasing difficulty. The duration of this test is about 15 minutes. EEG was not recorded while participants completed this test.

One-back task

Stimuli. The stimuli used in this study consisted of 49 images of faces, animals (e.g., giraffe, monkey, puppy), plants, objects (e.g., car, computer monitor, flower, banana), and scenes (e.g., city landscape, kitchen, bedroom). The 24 faces (13 identities, 8 males, and 8 neutral, 8 happy, 8 fearful expressions) were sampled from the Radboud Face dataset (Langner et al., 2010). The main facial features were aligned across faces using Procrustes transformations. Each face image was revealed through an ellipsoid mask that excluded non-facial cues. The non-face images were sampled from the stimulus set of Kiani et al. (Kiani et al., 2007). All stimuli were converted to 250 x 250 pixels (8x8 deg of visual angle) greyscale images. The mean luminance and the

luminance standard deviation of these stimuli were equalised using the SHINE toolbox (Willenbockel et al., 2010).

Procedure. We measured high-density electroencephalographic (EEG; sampling rate = 1024 Hz; 128-channel BioSemi ActiveTwo headset) activity while participants performed ~3200 trials of a one-back task in two recording sessions separated by at least one day and by a maximum of two weeks (**Fig. 5.1b**). Participants were asked to press a computer keyboard key on trials where the current image was identical to the previous one. Repetitions occurred with a 0.1 probability. They were asked to respond as quickly and accurately as possible. Feedback about accuracy was given on each trial. A trial unravelled as follows: a white fixation point was presented on a grey background for 500 ms (with a jitter of ± 50 ms); followed by a stimulus presented on a grey background for 600 ms; and, finally, by a white fixation point on a grey background for 500 ms. Participants had a maximum of 1100 ms following stimulus onset to respond. This interval, as well as the 200 ms preceding stimulus onset, constituted the epoch selected for our EEG analyses. In total, our participants completed 105,600 one-back trials which constituted ~32 hours of EEG epochs.

Analyses

All reported analyses were performed independently for each EEG recording session and then averaged. Analyses were completed using custom code written in MATLAB (MathWorks) and Python.

EEG preprocessing

EEG data was preprocessed using FieldTrip (Oostenveld et al., 2011): continuous raw data was first re-referenced relative to Cz, filtered with a band-pass filter [0.01-80 Hz], segmented into trial epochs from -200 ms to 1100 ms relative to stimulus onset, and down-sampled at 256 Hz.

Decoding analyses

Whole-brain analyses. To predict group-membership from EEG brain activity, we trained Fisher linear discriminant classifiers (5-fold cross-validation, 5 repetitions; Treder 2020; Grootswagers et al. 2017), using all 128 channels single-trial EEG data as features. Separate analyses were done on either all time points after stimulus onset, or successive 4-ms EEG time intervals. The classifiers were trained on trials of EEG activity of participants viewing faces and non-faces independently. The Area Under the Curve (AUC) was used to assess sensitivity. Additional control decoding analyses investigating effects of one-back trials on the predictions are shown in **Supplementary Fig. 5.1.**

Searchlight analysis. We conducted a searchlight analysis decoding EEG signals from all subsets of five neighbouring channels to characterise the scalp topographies of group-membership AUC. This searchlight analysis was done either using the entire EEG time series of a trial (0-1100 ms; **Fig. 5.2b,** leftmost topographies), or using 60 ms temporal windows (centred on 135 ms, 350 ms, 560 ms, and 775 ms; **Fig. 5.2b** rightmost topographies). We ran additional control searchlight decoding procedures investigating the effect of one-back trials (**Supplementary Fig. 5.1**).

Regression analysis. We used fractional ridge regression models (Rokem & Kay, 2020) to predict individual face recognition ability scores (CFMT+) among the typical recognisers from EEG patterns across time. We trained our model on subsets of 60% of the EEG patterns. We chose the alpha hyperparameter with the best coefficient of determination among 20 alpha hyperparameters ranging linearly from 0.001 to 0.99 applied on a 30% validation set. The decoding performance was assessed using the Spearman correlation between the CFMT+ scores and predictions from the overall best model (applied on the remaining 10% of EEG patterns). This process was repeated 10 times and the Spearman correlations were averaged. Significance was assessed using a permutation test (see *Group comparisons and inferential statistics* section).

Representational Similarity Analysis of brain and computational models

We compared our participants' brain representations to those from visual and semantic (caption-level) artificial neural networks using Representational Similarity Analysis (RSA; Charest et al., 2014; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008; Kriegeskorte & Kievit, 8/2013).

Brain Representational Dissimilarity Matrices. For every participant, we trained a Fisher linear discriminant to distinguish pairs of stimuli from every 4-ms intervals of EEG response (on all 128 channels) to these stimuli from -200 to 1100 ms after stimulus onset (Cichy & Oliva, 2020; Graumann et al., 2022). Cross-validated AUC served as pairwise classification dissimilarity metric. By repeating this process for all possible pairs (1176 for our 49 stimuli), we obtained a representational dissimilarity matrix (RDM). RDMs are shown for selected time points in **Supplementary Fig. 5.2**.

Visual Convolutional Neural Networks RDMs. We used a pre-trained AlexNet (Krizhevsky et al., 2012) as one model of the visual computations along the ventral stream (Güçlü & van Gerven, 2015). Our 49 stimuli were input to AlexNet. Layer-wise RDMs were constructed comparing the unit activation patterns for each pair of images using Pearson correlations. Similarly, we computed layer-wise RDMs from another well-known CNN, VGG-16 (see **Supplementary Fig. 5.3b**). Following previous studies using this model (Liu et al. 2021; Xie et al. 2020), we averaged the convolutional layer RDMs situated between each max pooling layers and the layers' input into five aggregated convolutional RDMs (e.g. conv1-1 & conv1-2 into RDM-conv1); this facilitated the comparison of our results with the five convolutional layers of AlexNet.

Semantic Caption-level Deep Averaging Neural Network RDM. We asked five new participants to provide a sentence caption describing each stimulus (e.g., "a city seen from the other side of

the forest", see **Fig. 5.1d**) using the Meadows online platform (www.meadows-research.com). The sentence captions were input in Google's universal sentence encoder (GUSE; Cer et al., 2018) resulting in 512 dimensional sentence embeddings for each stimulus. We then computed the dissimilarities (cosine distances) between the sentence embeddings across all pairs of captions, resulting in a semantic caption-level RDM for each participant. The average RDM was used for further analyses.

Comparing brain representations with computational models

We compared our participants' brain RDMs to those from the vision (Fig. 5.3a) and semantic (Fig. 3b) models described in the previous section using Conditional Mutual Information (CMI, Ince et al., 2017), which measures the statistical dependence between two variables (e.g. mutual information $I(x;y)$), removing the effect from a third variable (i.e. $I(x;y|z)$). Additional control comparisons using unconstrained Mutual Information between brain RDMs and both models are shown in **Supplementary Fig. 5.3a.**

Group comparison and inferential statistics

Comparison of Conditional Mutual Information time courses. Time courses of CMI were compared between the super-recognisers and typical recognisers using independent samples t-tests and a Monte Carlo procedure at a *p-value* of .05, as implemented in the Fieldtrip Toolbox (Oostenveld et al., 2011). Family-wise errors were controlled for using cluster-based corrections, with maximum cluster size as cluster-level statistic and an arbitrary *t* threshold for cluster statistic of [-1.90, 1.90] for the comparison of brain and semantic (excluding CNN) and [-2.75 2.75] for the comparison of brain and CNN (excluding semantic) time courses. The standard error is shown for all curves as colour-shaded areas (**Fig. 5.3**). Analyses with MI (brain; CNN) and MI (brain; semantic) were completed in an identical manner (**Supplementary Fig. 5.3**).

Time course of group-membership decoding. Significance was assessed using non-parametric permutation tests. We simulated the null hypothesis by training the linear classifier to identify shuffled group-membership labels from the experimental EEG patterns. This process was repeated 1000 times for each time point and each one of the two sessions. We then compared the real, experimental decoding value at each time point to its corresponding null distribution, and rejected the null hypothesis if the decoding value was greater than the prescribed critical value at a p <.001 level.

Time course of individual ability decoding using ridge regression. Significance was again assessed using non-parametric permutation testing. The ridge regression analysis predicted cross-validated CFMT+ scores from single trial EEG patterns, and goodness of fit is reported using Spearman's correlation between the predicted and observed CFMT+ scores. Under the null hypothesis that all participants elicited comparable EEG response patterns, irrespective of their CFMT+ score, the face recognition ability scores are exchangeable. We simulated this null hypothesis by repeating the ridge regression model training using randomly shuffled CFMT+ scores. The predicted CFMT+ scores were then correlated to the empirical, observed CFMT+ scores using Spearman's correlation, and this was repeated 1000 times for each time point. We finally compared the real, experimental correlation value with its corresponding null distribution at each time point, and rejected the null hypothesis if the correlation value was greater than the prescribed critical value at a p <.01 level.

Data availability

Data associated with this article will be available online upon publication of the manuscript.

Code availability

The MATLAB and Python codes used in this study will be available online upon publication of the manuscript.

Acknowledgements

We thank Prof. Josh P. Davis for sharing behavioural scores of super-recognisers and establishing first contact to the UK-based Super-Recognizers reported here. Funding for this project was supported by an ERC Starting Grant [ERC-StG-759432] to I.C, an ERSC-IAA grant to J.W., I.C. and S.F.S., by a Swiss National Science Foundation PRIMA (Promoting Women in Academia) grant [PR00P1_179872] to MR, and by NSERC and IVADO graduate scholarships to S.F.S. We also thank Mick Neville, from Super-Recognisers Ltd., who helped us to get in contact with some of our super-recognizer participants.

Author contributions

(CRediT standardised author statement)

S.F-S.: Conceptualisation, methodology, software, formal analysis, investigation, data curation, writing - original draft, visualisation, supervision, project administration, funding acquisition. **M.R.:** Investigation, resources, project administration, writing - review and editing. **E.B.:** investigation, project administration. **M.Z.**: investigation. **J.W.:** funding acquisition, writing - review and editing. **A-R.R.**: Investigation. **R.C.:** Resources. **F.G.:** Methodology, writing - original draft, supervision, funding acquisition. **I.C.**: Supervision, methodology, software, resources, formal analysis, writing - original draft, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

References

Arrington, M., Elbich, D., Dai, J., Duchaine, B., & Suzanne Scherf, K. (2022). Introducing the female Cambridge face memory test – long form (F-CFMT). In *Behavior Research*

Methods. https://doi.org/10.3758/s13428-022-01805-8

Anderson, A. J., McDermott, K., Rooks, B., Heffner, K. L., Dodell-Feder, D., & Lin, F. V. (2020). Decoding individual identity from brain activity elicited in imagining common experiences. *Nature Communications*, *11*(1), 5916. https://doi.org/10.1038/s41467-020-19630-y

Avidan, G., Hasson, U., Malach, R., & Behrmann, M. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 2. Functional Neuroimaging Findings. In *Journal of Cognitive Neuroscience* (Vol. 17, Issue 7, pp. 1150–1167). https://doi.org/10.1162/0898929054475145

Barton, J. J. S., Albonico, A., Susilo, T., Duchaine, B., & Corrow, S. L. (2019). Object recognition in acquired and developmental prosopagnosia. In *Cognitive Neuropsychology* (Vol. 36, Issues 1-2, pp. 54–84). https://doi.org/10.1080/02643294.2019.1593821

Barton, J. J. S., & Corrow, S. L. (2016). The problem of being bad at faces. *Neuropsychologia*, *89*, 119–124. https://doi.org/10.1016/j.neuropsychologia.2016.06.008

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences*, *9*(4), 180–187. https://doi.org/10.1016/j.tics.2005.02.011

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences*, *17*(5), 210–219. https://doi.org/10.1016/j.tics.2013.03.007

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*.

https://doi.org/10.1162/jocn.1996.8.6.551

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. In *Cortex* (Vol. 82, pp. 48–62). https://doi.org/10.1016/j.cortex.2016.05.003

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-Movement Strategies in Developmental Prosopagnosia and "Super" Face Recognition. In *Quarterly Journal of Experimental Psychology* (Vol. 70, Issue 2, pp. 201–217). https://doi.org/10.1080/17470218.2016.1161059

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology* , *77 (Pt 3)*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–1. https://doi.org/10.1167/13.10.1

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1803.11175

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570.

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating Predictors of Superior Face Recognition Ability in Police Super-recognisers: Superior face recognisers. *Applied Cognitive Psychology*, *30*(6), 827–840. https://doi.org/10.1002/acp.3260

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714–747. https://doi.org/10.1080/02643290500441296

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419–430. https://doi.org/10.1080/02643290701380491

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and …. *Neuropsychologia*. https://www.sciencedirect.com/science/article/pii/S0028393205002496

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*, 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Dwivedi, K., Bonner, M. F., Cichy, R. M., & Roig, G. (2021). Unveiling functions of the visual cortex using task-specific deep neural networks. *PLoS Computational Biology*, *17*(8),

e1009267. https://doi.org/10.1371/journal.pcbi.1009267

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain: A Journal of Neurology*, *135*(Pt 2), 542–554. https://doi.org/10.1093/brain/awr347

Faghel-Soubeyrand, S., Alink, A., Bamps, E., Gervais, R.-M., Gosselin, F., & Charest, I. (2019). The two-faces of recognition ability: better face recognizers extract different physical content from left and right sides of face stimuli. *Journal of Vision*, *19*(10), 136d – 136d. https://doi.org/10.1167/19.10.136d

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(6). https://doi.org/10.1073/pnas.2108091119

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience*, *23*(7), 1723–1740. https://doi.org/10.1162/jocn.2010.21545

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233. https://doi.org/10.1098/rsos.200233

Gabay, Y., Dundas, E., Plaut, D., & Behrmann, M. (2017). Atypical perceptual processing of faces in developmental dyslexia. *Brain and Language*, *173*, 41–51. https://doi.org/10.1016/j.bandl.2017.06.004

Garrido, L., Duchaine, B., & DeGutis, J. (2018). Association vs dissociation and setting appropriate criteria for object agnosia [Review of *Association vs dissociation and setting appropriate criteria for object agnosia*]. *Cognitive Neuropsychology*, *35*(1-2), 55–58. https://doi.org/10.1080/02643294.2018.1431875

Geskin, J., & Behrmann, M. (2018). Congenital prosopagnosia without object agnosia? A literature review. *Cognitive Neuropsychology*, *35*(1-2), 4–54. https://doi.org/10.1080/02643294.2017.1392295

Graumann, M., Ciuffi, C., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). The spatiotemporal neural dynamics of object location representations in the human brain. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01302-0

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562. https://doi.org/10.1038/nn1224

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Harel, A., Kravitz, D., & Baker, C. I. (2013). Beyond perceptual expertise: revisiting the neural substrates of expert object recognition. *Frontiers in Human Neuroscience*, *7*, 885. https://doi.org/10.3389/fnhum.2013.00885

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The

representational dynamics of task and object processing in humans. In *eLife* (Vol. 7). https://doi.org/10.7554/elife.32816

Hendel, R. K., Starrfelt, R., & Gerlach, C. (2019). The good, the bad, and the average: Characterizing the relationship between face and object processing across the face recognition spectrum. *Neuropsychologia*, *124*, 274–284. https://doi.org/10.1016/j.neuropsychologia.2018.11.016

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573. https://doi.org/10.1002/hbm.23471

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), E6418–E6427. https://doi.org/10.1073/pnas.1802246115

Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 861–878. https://doi.org/10.3758/s13415-013-0234-y

Kaneshiro, B., Perreau Guimaraes, M., Kim, H.-S., Norcia, A. M., & Suppes, P. (2015). A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. *PloS One*, *10*(8), e0135697.

https://doi.org/10.1371/journal.pone.0135697

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309.

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Kriegeskorte, N., & Kievit, R. A. (8/2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lohse, M., Garrido, L., Driver, J., Dolan, R. J., Duchaine, B. C., & Furl, N. (2016). Effective Connectivity from Early Visual Cortex to Posterior Occipitotemporal Face Areas Supports Face Selectivity and Predicts Developmental Prosopagnosia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(13), 3821–3828. https://doi.org/10.1523/JNEUROSCI.3621-15.2016

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38), E9015–E9024. https://doi.org/10.1073/pnas.1719616115

Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: the brains behind superior memory. *Nature Neuroscience*, *6*(1), 90–95. https://doi.org/10.1038/nn988

McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the*. https://www.pnas.org/content/99/3/1645.short

Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: repeat assessment using the Cambridge Face Memory Test. *Royal Society Open Science*, *7*(9), 200884. https://doi.org/10.1098/rsos.200884

Naselaris, T., Allen, E., & Kay, K. (2021). Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences*, *40*, 45–51. https://doi.org/10.1016/j.cobeha.2020.12.008

Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science*, *8*(3), 201169. https://doi.org/10.1098/rsos.201169

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. In *Nature Neuroscience* (Vol. 24, Issue 11, pp. 1628–1636).

https://doi.org/10.1038/s41593-021-00921-6

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421. https://doi.org/10.1016/s1364-6613(02)01976-9

Ramon, M. (2021). Super-Recognizers –a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, 107809. https://doi.org/10.1016/j.neuropsychologia.2021.107809

Rokem, A., & Kay, K. (2020). Fractional ridge regression: a fast, interpretable reparameterization of ridge regression. *GigaScience*, *9*(12). https://doi.org/10.1093/gigascience/giaa133

Rosenthal, G., Tanzer, M., Simony, E., Hasson, U., Behrmann, M., & Avidan, G. (2017). Altered topology of neural circuits in congenital prosopagnosia. *eLife*, *6*. https://doi.org/10.7554/eLife.25069

Rossion, B., & Jacques, C. (2012). The N170: Understanding the time course of face perception in the human brain. *The Oxford Handbook of Event-Related Potential Components.*, *641*, 115–141. https://psycnet.apa.org/fulltext/2013-01016-005.pdf

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Shen, W., Fiori-Duharcourt, N., & Isel, F. (2016). Functional significance of the semantic P600: evidence from the event-related brain potential source localization. *Neuroreport*, *27*(7), 548–558. https://doi.org/10.1097/WNR.0000000000000583

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale

Image Recognition. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1409.1556

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science*, *30*(2), 300–308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338

Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Research. Cognitive Brain Research*, *22*(2), 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

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

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*, *29*(5-6), 360–392. https://doi.org/10.1080/02643294.2012.753433

Zadelaar, J. N., Weeda, W. D., Waldorp, L. J., Van Duijvenvoorde, A. C. K., Blankenstein, N. E., & Huizenga, H. M. (2019). Are individual differences quantitative or qualitative? An integrated behavioral and fMRI MIMIC approach. *NeuroImage*, *202*, 116058. https://doi.org/10.1016/j.neuroimage.2019.116058

Supplementary material

Behavioural results

All participants' face recognition ability was assessed using the Cambridge Face Memory Test long-form (CFMT+, (Russell et al. 2009). Scores on the CFMT+ ranged from 50 to 85 in the typical recognisers group (M_{TRs}=70.00; SD=9.09), and from 92 to 100 in the experimental super-recogniser group (M_{SRs}=95.38, SD=2.68; difference between groups : $t(31)$=10.6958, p<.00001 see Fig. 1a). The main experimental task was a one-back task (Fig. 1b). Accuracy was significantly greater for the super-recognisers (M_{SRs}=.93, SD=.054) than for the typical recognisers (M_{TRs}=.83, SD=.094; $t(31)$=3.7911, p=0.00065). This was also true when analysing separately face (M_{SRs}= .9260, SD=.0471; M_{TRs}=.8144, SD=.1066; $t(31)$=3.8440, p=0.00056) and non-face trials (M_{SRs}=.9456, SD=.0651; M_{TRs}=.8532, SD=.0929; $t(31)$=3.2855, p=.0025). Furthermore, accuracy in the one-back task was positively correlated with scores on the CFMT+ (r=.68, p<.001; RT was marginally associated with CFMT+, r=.37, p=.04). We observed no significant differences in response times between the two groups (p>.3).

Scores obtained by the super-recogniser tested in the UK on a battery of face recognition tests

subject-ID	CFMT+	GFMT	Face Array	LASIE Match	Black and White	Super-rec	Longterm	one-back faces	one-back non-faces
SR-1	93	40	38	83	34	13	9	.9328	0.977
SR-2	95	40	37	88	33	12	9	.8500	0.7903
SR-3	97	40	32	84	33	12	8	.9325	0.9718
SR-4	93	40	38	92	34	13	8	.9726	0.9835
SR-5	98	40	31	91	33	14	9	.9326	0.9787
SR-6	100	40	39	82	40	14	10	.9823	0.9817
SR-7	98	40	40	92	33	13	8	.9319	0.9647
SR-8	96	38	40	90	39	12	10	.9362	0.9787
Max score	102	40	40	100	40	14	10	1	1

Supplementary table 1: The scores above show performance on a battery of standardised face recognition tests (Noyes et al. 2021) for the participants identified as super-recognisers in the UK. Also shown are the scores for our one-back task, for face and non-face trials. The last row of the table shows the maximum obtainable score for each test.

Scores obtained by the super-recogniser tested in Switzerland on a battery of face recognition tests

subject-ID	CFMT+	FICST score	YBT long raw score	one-back faces	one-back non-faces
SR-9	92	0	29	.8132	0.8384
SR-10	99	0	17	.9409	0.9619
SR-11	92	0	20	.9635	0.9808
SR-12	93	1	20	.8581	0.8212
SR-13	92	7	17	.9530	0.985
SR-14	96	0	18	.9604	0.9886
SR-15	94	3	16	.9112	0.9624
SR-16	97	1	22	.9792	0.9643
Best score	102	0	35	1	1

Supplementary table 2: The scores above show performance on a battery of standardised face recognition tests (Ramon, 2021) for the participants identified as super-recognisers in Switzerland. Also shown are the scores for our one-back task, for face and non-face trials. The last row of the table shows the best obtainable score for each test.

Noise ceiling for group-membership decoding using CFMT+ reliability scores

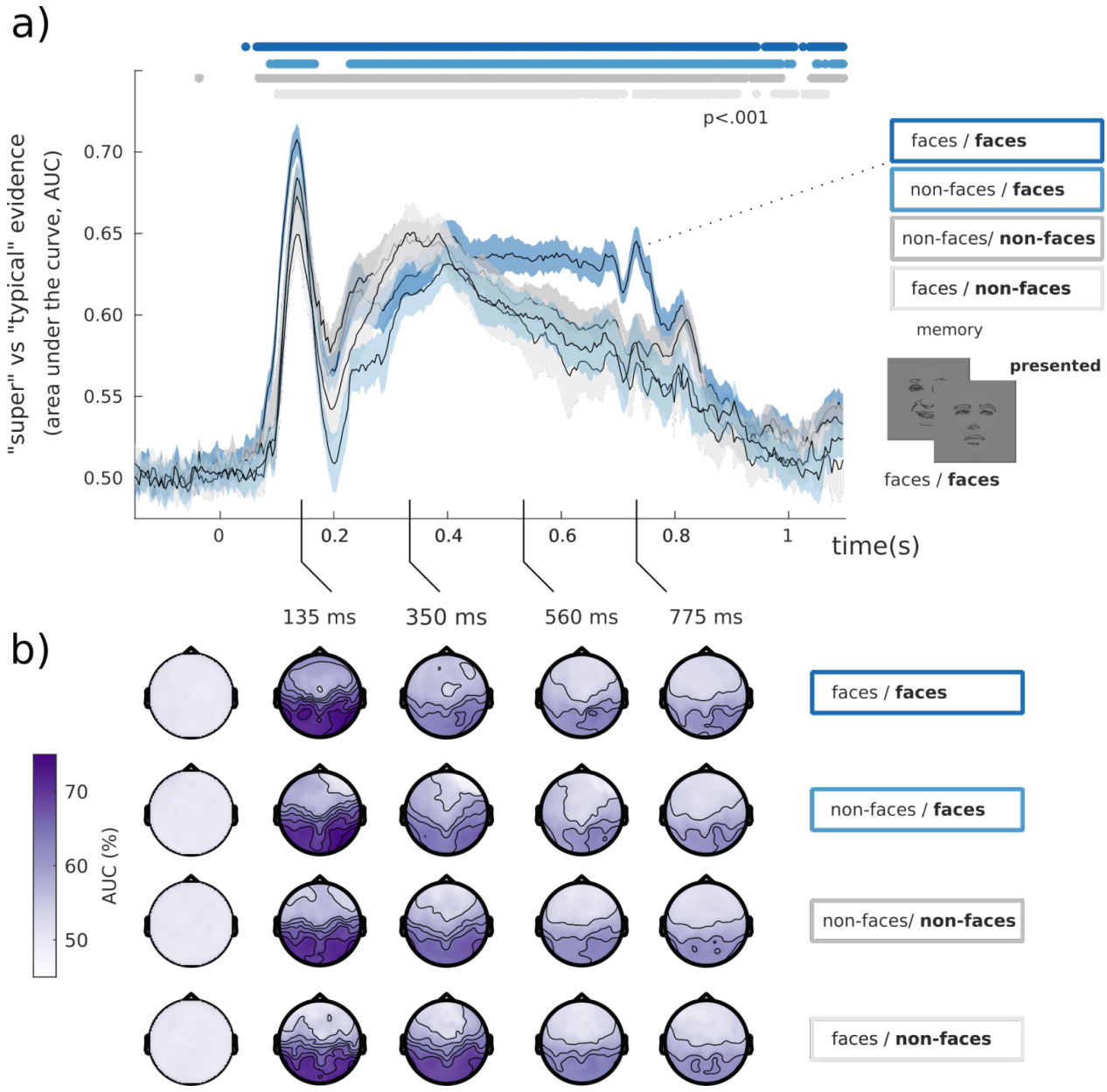
To better interpret the magnitude of our group-membership decoding accuracies, we determined the maximum attainable accuracy when categorising a super-recogniser as such using the CFMT+ (i.e., the empirically imposed noise ceiling for our decoding group-membership analysis). We simulated 1000 distributions of CFMT+ scores, each with N=32, and each having a Pearson correlation of .71±.01 with the distribution of CFMT+ scores observed in this study. The correlation coefficient of .71 was chosen according to the test-retest reliability the CFMT+ measured in previous studies (Arrington et al., 2022; Murray & Bate, 2020). We then predicted super-recogniser participants from these distributions: we compared the simulated SR label (i.e. simulated CFMT+ scores higher than the prescribed cut-off of 92) to the used SR labels. We then

averaged the accuracies across the simulated participants, and computed the maximum accuracy from these 1000 distributions. This process was repeated 10000 times, creating a distribution of 10000 simulated maximums with a mean of M=0.9310 (SD=0.0228) that we interpreted as the noise ceiling.

Univariate analyses of EEG associated with individual ability

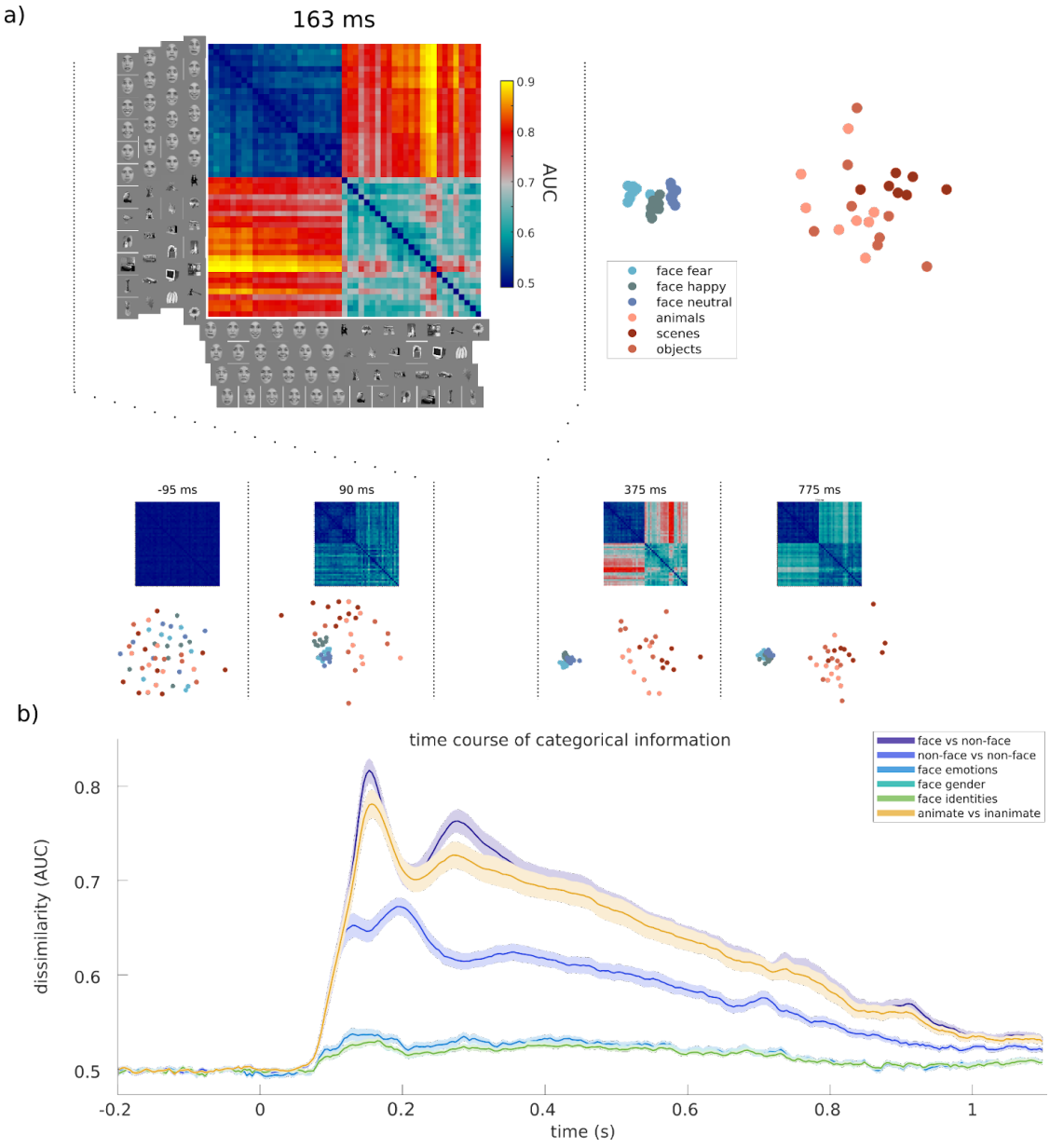
N170 amplitude and latency. We also performed more traditional event-related potential (ERP) analyses for both groups. We extracted, for face and non-face trials, peak negative ERP amplitudes and latencies for every participant in a window corresponding to the N170 component (within 110-200 ms at electrodes [B6, B7, B8, A28] on the right hemisphere and [A9, A10, A11, A15] on the left hemisphere). We tested the conditions, groups, and their interaction effects using an ANOVA on N170 peak latency and amplitude separately. No interaction effects were observed for peak latencies ($F_{interaction}(60,1)=1$, $p=.32$) and amplitudes ($F_{interaction}(60,1)=0.32$, $p>.50$). The peak N170 was earlier ($F_{conditions}(60,1)=5.86$, $p=.0185$) and presented greater amplitudes ($F_{conditions}(60,1)=33.78$, $p<.0001$) for faces compared to non-face objects. Moreover, the peak N170 was earlier ($F_{group}(60,1)=19.23$, $p<.0001$) and presented larger amplitudes ($F_{group}(60,1)=13.75$, $p=.0005$) in super-recognisers than typical recognisers.

Lateralisation. Compared to typical recognisers, super-recognisers showed greater N170 peak amplitudes in the right-hemisphere electrodes for faces (computed as the difference between the right [B6, B7, B8, A28] and left [A9, A10, A11, A15]; $t(30)=-2.8542$, $p=.01$). Moreover, the CFMT+ scores of typical recognisers correlated with right-hemisphere lateralisation of the N170 for faces ($r(16)=-.53$, $p=.0298$). These effects were not significant either for non-face stimuli, or for N170 peak latency.



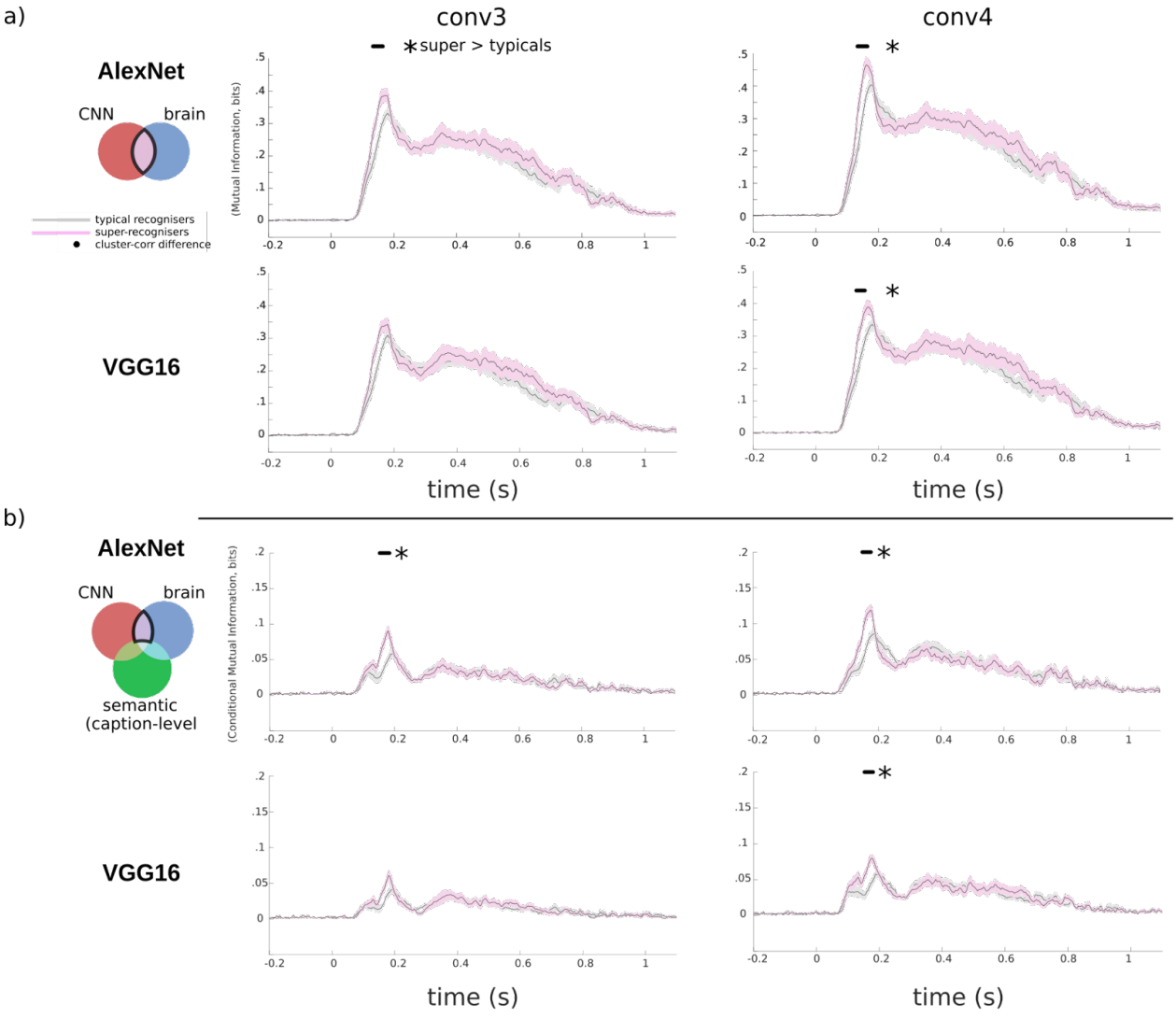
Supplementary figure 5.1. **a)** We computed the time course of decoding accuracy for group membership from all different face and non-face combinations of one-back and current trials (e.g. consecutive face - face trials). We observed a similar time course for all combinations. However the consecutive face-face trials showed larger decoding accuracies around 400 to 750 ms. **b)** The topographies show results from a searchlight decoding analysis with classification performance attaining 75% accuracy around 135 ms over occipito-temporal electrodes for face presented conditions (74.6% for face-face, 75.1% for nonface-face) and 72% for non-face presented conditions (71.7% for nonface-nonface, 71.5% for face-non-face). Note that drawings of faces

are depicted here as an anonymised substitute to the experimental face stimuli presented to our participants.



Supplementary figure 5.2. EEG representational geometry dynamics. a) Representational Similarity Analysis (RSA) was applied to time-resolved EEG patterns, using decoding AUC as dissimilarity measure between pairs of images (5 fold cross-validation, 5 repetitions) to create Representational Dissimilarity Matrices (RDMs). Multidimensional scaling was employed to

visualise these high-dimensional brain representations on a 2D plane, which showed clear distinctions between various categories (e.g. face clusters, scenes clusters, animal clusters, etc.). b) We revealed categorical information unfolding in time by averaging dissimilarities between stimulus categories (e.g. faces vs non-face objects) and averaging across participants. Brain representations for the distinction of face vs. non-face objects (a hallmark of the N170, (Rossion & Jacques, 2012) dominated all other categorical distinctions (Carlson et al., 2013; Kaneshiro et al., 2015), and peaked at 153 ms.



Supplementary figure 5.3. Comparison of super- and typical-recogniser brain representations with those of artificial neural networks of vision. a) Mutual information results comparing brain RDMs and AlexNet RDMs (first row) and brain RDMs and VGG16 RDMs (second row) are shown for typical- (grey curve) and super-recognisers (pink curve). We

found greater similarity with mid-level visual computations as indexed from both CNN models (layers 3, 4 shown for AlexNet and VGG 16) in the brains of super-recognisers (black line indicates significant contrasts, $p<.05$, cluster-corrected) between 130 ms to 160 ms. **b**) We also computed the MI between CNN RDMs and the brain RDMs, but constrained on the Mutual Information from the caption-level semantic RDM. Again, we found greater similarity with mid-level visual computations as indexed from both CNN models (layers 3, 4 shown for AlexNet and VGG 16) in the brains of super-recognisers *(*$p<.05$, cluster-corrected) between 133 ms to 165 ms. The observed magnitudes were reduced as expected given the additional constraints. The shaded areas of all curves represent the standard error.

Chapitre 6 – Article 5

Neural computations in prosopagnosia

Simon Faghel-Soubeyrand [1,2], Anne-Raphaelle Richoz [3], Delphine Waeber [3], Jessica Woodhams [2], Frédéric Gosselin[1], Roberto Caldara [3] & Ian Charest[1,2]

1. Département de psychologie, Université de Montréal, Canada
2. School of Psychology and Centre for Human Brain Health, University of Birmingham, Birmingham, UK
3. Department of Psychology, University de Fribourg, Switzerland
4. School of Psychology, University of Birmingham, Birmingham, UK

Manuscript currently under review at *Cerebral Cortex*.

Highlights

- We assess the neural computations in the prosopagnosic patient PS using EEG, RSA, and deep neural networks
- Neural dynamics of brain-lesioned PS are reliably captured using RSA
- Neural decoding shows normal evidence for non-face individuation in PS
- Neural decoding shows abnormal neural evidence for face individuation in PS
- PS shows impaired high-level visual and semantic neural computations

Abstract

We aimed to identify neural computations underlying the loss of face identification ability by modelling the brain activity of brain-lesioned patient PS, a well-documented case of acquired pure prosopagnosia. We collected a large dataset of high-density electrophysiological (EEG) recordings from PS and neurotypicals while they completed a one-back task on a stream of face, object, animal and scene images. We found reduced neural decoding of face identity around the N170 window in PS, and conjointly revealed normal *non-face* identification in this patient. We used Representational Similarity Analysis (RSA) to correlate human EEG representations with those of deep neural network (DNN) models of vision and caption-level semantics, offering a window into the neural computations at play in patient PS's deficits. Brain representational dissimilarity matrices (RDMs) were computed for each participant at 4 ms steps using cross-validated classifiers. PS's brain RDMs showed significant reliability across sessions, indicating meaningful measurements of brain representations with RSA even in the presence of significant lesions. Crucially, computational analyses were able to reveal PS's representational deficits in high-level visual and semantic brain computations. Such multi-modal data-driven characterisations of prosopagnosia highlight the complex nature of processes contributing to face recognition in the human brain.

Word counts. Abstract: 203. Main text : 4501, Method : 2159

Key words : prosopagnosia, EEG, RSA, artificial neural networks, brain dynamics, semantic representations, visual representations

Introduction

The human brain is equipped with sophisticated machinery optimised to quickly and effectively recognise faces in a series of computations unfolding within tens of milliseconds. A dramatic contrast to this typically efficient process has been revealed in brain-lesioned patients with an inability to recognise faces, individuals called acquired prosopagnosics (Bodamer, 1947). Findings from these patients have refined the functional role and the distributed nature of the face-sensitive brain regions in the ventral stream, such as the fusiform gyrus (FFA; Bobes et al., 2003; Kanwisher et al., 1997) and the lateral portion of the inferior occipital gyrus (Occipital Face Area, OFA; Dricot et al., 2008; Gauthier et al., 2000; Rossion et al., 2003; Sorger et al., 2007). This literature has generally contributed to the idea that specialised and category-selective neural modules are necessary for functional aspects of face processing (Cohen et al., 2019). Brain imaging findings from individuals born with deficits in face recognition (developmental prosopagnosics; (Avidan et al., 2005; Jiahui et al., 2018; Kaltwasser et al., 2014; McConachie, 1976; Rosenthal et al., 2017) have revealed finer-grained functional neural differences in the processes (Avidan et al., 2014; Jiahui et al., 2018; Rosenthal et al., 2017) associated with deficits in face recognition. Overall, the cumulation of these neuropsychological, neuroanatomical and functional components of prosopagnosia (Busigny et al. 2010; Dricot et al. 2008; Rossion 2018; Rossion et al. 2003; Duchaine and Nakayama 2006) has significantly contributed to neural models of face perception in the last two decades (Duchaine & Yovel, 2015; Haxby et al., 2000; White & Mike Burton, 2022). Yet, very little is known on the nature of face representations of those patients (e.g., Caldara et al., 2005; Fiset et al., 2017), and next to nothing is known on the nature of brain dynamics and neural computations affected in prosopagnosia. Here, we report an investigation of the neural computations involved in the processing of faces and objects of patient PS, a well-documented case of pure acquired prosopagnosia (Rossion et al., 2003; Sorger et al., 2007), using Representational Similarity Analysis (RSA) applied to brain imaging and

computational models.

Patient PS is a right-handed woman having sustained a closed head injury in 1992, leading to extensive bilateral occipitotemporal lesions encompassing the right OFA, left FFA, and a small region of the right middle temporal gyrus (Dricot et al., 2008; Sorger et al., 2007). She is perhaps the most studied case of acquired prosopagnosia, with more than 32 scientific publications in the last 20 years (see Rossion, 2022a, 2022b for recent reviews on this patient). The particular attention given to this case can be attributed to the relatively focal aspect of her lesions in the face network and the resulting highly specific impairment for face-identification (Busigny et al., 2010). The neuro-anatomical/functional basis of PS has already been exhaustively reviewed elsewhere (Rossion, 2022b). Overall, while her condition has been shown to affect a wide array of perceptual mechanisms (e.g. holistic processes, Ramon et al., 2016; the visual content of face representations in Caldara et al., 2005; Fiset et al., 2017; see also Rossion, 2022a), a direct characterisation of the neural computations behind her deficits has, to the best of our knowledge, never been attempted. A traditional proxy to the nature and level of brain computations affected in this patient, and in prosopagnosia in general, has been to consider the temporal dynamics and face-selectivity of the underlying neural activity. Event-related potential differences occurring late, for example, are generally interpreted as representing higher level processes than those occurring earlier (Alonso Prieto et al., 2011; Bentin & Deouell, 2000; Eimer et al., 2012; Gosling & Eimer, 2011; Herzmann et al., 2004; Liu-Shuang et al., 2016a; Simon et al., 2011; Tanaka et al., 2006; Wiese et al., 2019a). Some associations have been revealed between prosopagnosia and typical neural correlates of face-processing, like the face-sensitive N170 (Bentin et al., 1996) and face-selective fMRI activation (Towler and Eimer 2012; Bobes et al. 2003; Alonso Prieto et al. 2011; Gao et al. 2019). Interestingly, however, despite important lesions and behavioural deficits in face-identification, PS still shows typical face-selectivity in spared-regions of the right-hemisphere (i.e. she displays a right FFA; e.g. see (Gao et al., 2019; Rossion et al., 2003), as well as a typical N170 component in the right, but not left-hemisphere (Alonso-Prieto, 2011; see also Bobes et al., 2003; Dalrymple et al., 2011). Similarly, developmental prosopagnosics show

typical activation across the “core” (posterior) regions of the face-processing system (OFA and FFA, e.g. Avidan et al., 2014). More recently, robust experimental techniques such as fast periodic visual stimulation (Liu-Shuang et al., 2016b) have been able to shed light on the important deficits in neural face-individuation of PS. Characterising the underlying computations of these neural and perceptual processes, however, remains a challenging task. First, describing neural computations is generally arduous due to signal-to-noise ratio (SNR) issues, which is even more concerning when recording brain activity from brain-lesion patients (Liu-Shuang et al., 2016a). Brain damage, for example, can significantly alter the flow of brain activity compared to typical observers, which can potentially deform event-related potential components (Alonso Prieto et al., 2011) and require more repetitions of conditions. Second, using solely temporal evidence is limited in itself to reveal brain computations as it only partially indicates the nature of the computations that are relied on by the brain (Lamme & Roelfsema, 2000; McDermott et al., 2002). Individuals relying on different neural computations in responses to faces, for example, could have identical activity at a given latency as indexed by univariate event-related potentials.

More explicit ways of revealing the nature of brain representations have recently gained traction with techniques associating functional and multivariate brain activity to computational models (Dwivedi et al. 2021; di Oleggio Castello et al. 2021; Popham et al. 2021; Kriegeskorte and Diedrichsen 2016; Doerig et al. 2022; Faghel-Soubeyrand et al. 2022). The aforementioned SNR concerns might explain why most work on prosopagnosia has relied on a limited set of stimuli conditions, block-designs, and univariate methods such as averaging of conditions and subtraction approaches. However, while significantly improving the statistical power of these studies, these approaches have prevented a thorough description of the brain computations underlying prosopagnosia (Friston et al., 2006). Investigating brain processing using condition-rich designs (Allen et al., 2022; Charest et al., 2014a; Kriegeskorte & Kievit, 2013; Naselaris et al., 2021), and promoting a broad description of underlying brain mechanisms by testing diverse models on a whole-brain basis (Dwivedi et al., 2021; Kriegeskorte & Diedrichsen, 2019; Popham et al., 2021) might provide a more comprehensive understanding of these

processes.

Here, we take into account these temporal and computational aspects of brain processes by investigating prosopagnosia with fine-grained temporal recordings of brain activity (high-density electroencephalography; EEG), machine learning, and a proven multivariate method, i.e. Representational Similarity Analysis (Kriegeskorte, Mur, & Bandettini, 2008). We recorded the brain activity of PS and neurotypical controls in responses to images of various categories. Using multivariate pattern analysis (time-resolved "decoding"; Grootswagers et al., 2017), we probe PS' neural evidence for face and non-face identity representations. Using RSA, we produce functional brain representations in a format that provides straightforward comparisons between individuals differing in neuroanatomical structure (Golarai et al., 2015; Popal et al., 2019). This enabled us to compare human brain representations with those of artificial models characterising different types of computations, i.e. deep neural networks of vision and semantic classification, thereby offering a window into the neural computations at play in patient PS's deficits.

Materials and Procedures

Patient PS and neurotypical participants

A total of 20 participants were recruited for this study. The first group consisted of 19 neurotypicals individuals that included 15 healthy controls (9 female, Mage = 22.9 years old) as well as 4 aged-matched (3 female, Mage = 67.5). This sample size was chosen according to the effects described in previous mvpa studies (Carlson et al., 2013; Cichy et al., 2014; Faghel-Soubeyrand et al., 2022b; Hebart et al., 2018), as well as previous studies on prosopagnosia (Gao et al., 2019; Humphreys et al., 2007; Liu-Shuang et al., 2016a; Richoz et al., 2015). Data from 10 of these participants (healthy controls 1-10) have been reported in a previous study (Faghel-Soubeyrand et al., 2022b). One participant from the aged-matched group (aged-matched #2) was rejected due to faulty EEG recordings and poor behavioural performance during the one-back task and CFMT+. This study was approved by the Ethics and Research

Committee of the University of Birmingham, The University of Fribourg, and informed consent was obtained from all participants.

PS's case report

Patient PS was born in 1950 and is a *pure* case of acquired prosopagnosia. She was hit by the side mirror of a London's bus in 1992 while crossing the road. This closed head injury led to major lesions in the left middle fusiform gyrus, where the left Fusiform Face Area (lFFA) is typically located, and in the right inferior occipital gyrus, which typically locates the right Occipital Face Area (rOFA; see Gao et al. 2019 for converging fMRI evidence). Both regions play a critical functional role within the face cortical network (Rossion, 2022a, 2022b). She also reported minor damages in the right middle temporal gyrus and left posterior cerebellum (for an exhaustive anatomical description and an illustration of her brain damages see, Sorger et al., 2007, Figures 2 and 3). Patient PS is a very well-documented and described case of acquired prosopagnosia. She has been extensively studied over the last 20 years, leading to impactful scientific contributions that significantly enriched the theoretical models on human face perception (Rossion, 2008, 2014; for a complete case report see, Rossion, 2022a, 2022b; Rossion et al., 2003b).

Patient PS recovered remarkably well from initially significant cognitive deficits with the support of medical treatment and neuropsychological rehabilitation. A couple of months after her injury she performed within the normal range at different non-visual tasks for which she was slightly impaired after the accident (e.g., calculation, short and long-term memory, visual imagery). Yet, her fine-grained visual discrimination abilities remained slower compared to controls, and she also presented reduced contrast sensitivity to high spatial frequency information (>22c/degree) and a profound prosopagnosia with massively impaired face recognition abilities (Rossion et al., 2003). The patient complains of a severe difficulty at recognizing faces, including the ones of close relatives (husband, children, friends), as well as her own face. PS can correctly categorise (and draw) faces as a unique visual object and discriminate faces from other non-face objects or

scenes, even when the images are briefly presented (Schiltz et al., 2006). She shows no difficulty at object recognition, even for subordinate-level discriminations (Rossion et al., 2003; Schiltz et al., 2006). Patient PS is perfect at all tests from the Birmingham Object Recognition Battery (BORB – Riddoch & Humphreys, 2022) showing preserved processing of low-level aspects of visual information (i.e., matching of basic elementary features), intact object matching from different viewpoints, and normal performance for object naming (Rossion et al., 2003; Table 1). Her reading abilities are also well preserved although slightly slowed down, her visual acuity (0.8 bilaterally) is within the normal range, and her visual field almost intact apart from a small left paracentral scotoma. As reported by Rossion et al. (2003), she is highly impaired on the Benton Face Matching Test (BFRT - Benton & Van Allen, 1972) scoring 27/54 (percentile 1). She performs also poorly on the Warrington Recognition Memory Test (WRMT - Warrington & Shallice, 1984), scoring 18/25 (percentile 3) a performance that characterises her as impaired compared to controls. Over the years, patient PS developed strategies to infer a person's identity by relying on external cues such as haircut, clothes, beard, glasses, gait, posture, or a person's voice. Moreover, as revealed by the *Bubbles* response classification technique, patient PS uses suboptimal diagnostic information to recognise familiar faces, relying on the lower part of the face (i.e., the mouth region and external contours) instead of the most informative eye area (Caldara et al., 2005). A similar bias towards the mouth has been observed for the recognition of static facial expressions (Fiset et al., 2017) for which she is strongly impaired. Her ability to recognize the dynamic versions of the same facial expressions is nevertheless preserved (Richoz et al., 2015). Overall, PS is a very cooperative patient with extraordinarily preserved cognitive functions, sensory and motor skills and without any attentional deficits. She therefore represents an exemplary case to investigate the functional models of typical face processing.

Behavioural tasks

One-back task. The stimuli used in the main experiment consisted of 49 images of faces, animals (e.g., giraffe, monkey, puppy), plants, objects (e.g., car, computer monitor, flower,

banana), and scenes (e.g., city landscape, kitchen, bedroom). The 24 faces (13 identities, 8 males, and 8 neutral, 8 happy, 8 fearful expressions) were taken from the Radboud Face dataset (Langner et al., 2010). For further details on stimulus processing steps, see Faghel-Soubeyrand et al. (2022).

These stimuli were presented during a one-back task where we measured high-density electroencephalographic (EEG) activity (**Fig. 6.1b,c**). Participants performed ~3200 trials in two recording sessions, which were separated by at least one day and by a maximum of two weeks. Participants were asked to press a computer keyboard key on trials where the current image was identical to the previous one (repetitions occurring with a 0.1 probability). They were asked to respond as quickly and accurately as possible. Feedback about accuracy was given on each trial. A trial unravelled as follows: a white fixation point was presented on a grey background for 500 ms (with a jitter of ± 50 ms); followed by a stimulus presented on a grey background for 600 ms; and, finally, by a white fixation point on a grey background for 500 ms. Participants had a maximum of 1,100 ms following stimulus onset to respond. This interval, as well as the 200 ms preceding stimulus onset, constituted the epoch selected for our EEG analyses.

Cambridge Face Memory Test +. All participants were administered the CFMT long-form, or CFMT+ (Russell et al., 2009). In the CFMT+, participants are required to memorise a series of face identities, and to subsequently identify the newly learned faces among three faces. It includes a total of 102 trials of increasing difficulty. The duration of this test is about 15 minutes. EEG was not recorded while participants completed this test.

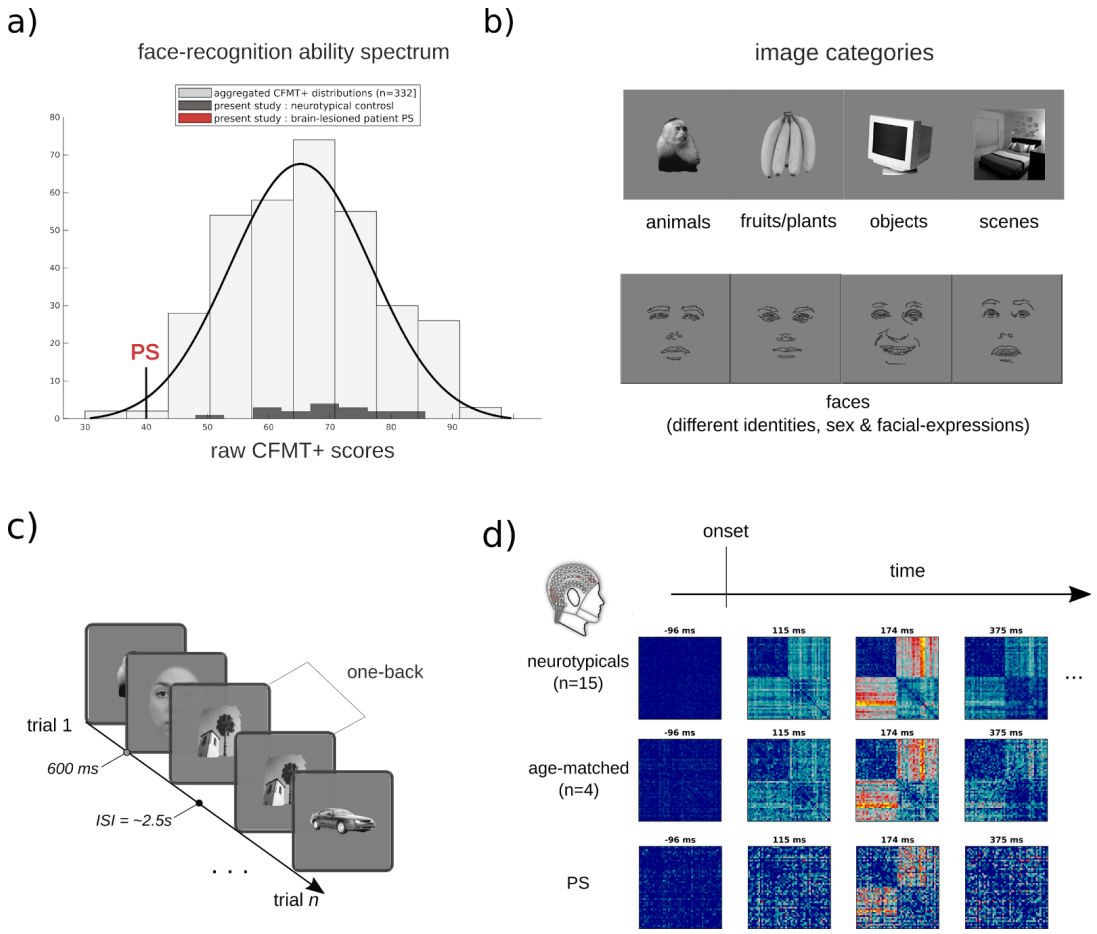


Figure 6.1. Overview of the experimental procedure. a) The histogram shows the Cambridge Face Memory Test long-form (CFMT+, Russell et al., 2009) scores of PS, typical recognisers (black bars), and an additional 332 neurotypical observers from three independent studies for comparison (Faghel-Soubeyrand et al., 2019; Fysh et al., 2020; Tardif et al., 2019). **b)** Participants engaged in a one-back task while their brain activity was recorded with high-density electroencephalography. The objects depicted in the stimuli belonged to various categories, such as faces, objects, and scenes. Note that the face drawings shown here are an anonymised substitute to the experimental face stimuli presented to our participants. **c)** Brain Representational dissimilarity matrices (RDM) were computed for PS and controls using cross-validated decoding performance between the EEG topographies from each pair of stimuli at every 4 ms time-point. Brain RDMs are shown for specific time points.

EEG recording and preprocessing

High-density electroencephalographic data was continuously recorded at a sampling rate of 1024 Hz using a 128-channel BioSemi ActiveTwo headset (Biosemi B.V., Amsterdam, Netherlands).

Electrodes' impedance was kept below 20 μV. Data were collected at the University of Fribourg. Data was preprocessed using FieldTrip (Oostenveld et al., 2011) and in-house Matlab code: continuous raw signal was first re-referenced relative to A1 (Cz), filtered with a band-pass filter [.01-80 Hz], segmented into trial epochs from -200 ms to 1100 ms relative to image onset, and down-sampled at 256 Hz. These EEG recordings were completed during the one-back task only.

Representational Similarity Analysis

We compared our participants' brain representations to those from visual and caption deep neural networks using Representational Similarity Analysis (RSA; (Charest et al., 2014b; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008; Kriegeskorte & Kievit, 8/2013).

Brain Representational Dissimilarity Matrices. For every participant, we trained a Fisher linear discriminant (5-fold cross-validation, 5 repetitions; Treder, 2020) to distinguish pairs of stimuli from every 4-ms intervals of EEG response to these stimuli from -200 to 1100 ms after stimulus onset (Cichy & Oliva, 2020; Graumann et al., 2022). All 128 channels served as features in these classifiers. Cross-validated area under the curve (AUC) served as pairwise classification dissimilarity metric. By repeating this process for all possible pairs (1176 for our 49 stimuli), we obtained a representational dissimilarity matrix (RDM). RDMs are shown for selected time points in **Figure 6.1d.**

Visual Convolutional Neural Networks RDMs. We used a pre-trained AlexNet (Krizhevsky et al., 2012) as one model of the visual computations along the ventral stream (Güçlü & van Gerven, 2015). Our 49 stimuli were input to AlexNet. Layer-wise RDMs were constructed comparing the unit activation patterns for each pair of images using Pearson correlations. This CNN process visual features of gradually higher complexity and abstraction along their layers (Güçlü & van Gerven, 2015), from low-level (i.e., orientation, edges in shallow layers) to high-level features (e.g., objects and object parts in deeper layers).

Caption-level Semantic RDM. To derive a model of an higher level of computations than purely visual processes, we first asked 5 new participants to provide a sentence caption describing each stimulus (e.g., "a city seen from the other side of the forest") using the Meadows online platform (www.meadows-research.com). The sentence captions were fed as input in Google's universal sentence encoder (GUSE; Cer et al., 2018) resulting in 512 dimensional sentence embeddings. GUSE has been trained to predict semantic textual similarity from human judgments, and its embeddings generalise to an array of other semantic judgement tasks (Cer et al., 2018). We then computed the dissimilarities (cosine distances) between the sentence embeddings across all pairs of captions, resulting in a caption-level semantic RDM for each participant. The average RDM was used for further analyses (see **Supplementary figure 6.2)**.

Decoding analyses

Face-identity decoding. We trained multiclass linear discriminant classifiers to predict 8 face identities (5-fold cross-validation, 5 repetitions; Grootswagers et al., 2017; Treder, 2020), using all 128 channels single-trial EEG data as features. A sliding-average with a window of 39 ms was applied to EEG traces prior to training. Separate classifiers were trained on the resulting successive 4-ms EEG time intervals. The classifiers were trained on trials of EEG activity of participants viewing face-identities varying in facial-expressions (fear & joy; ~525 observations per session per participant), similarly to previous studies of face-identity information using EEG signal (Nemrodov et al., 2016). The time courses of these decoders' performance were averaged across sessions. Recall (defined as : *true positives* / [*true positives* + *false negatives]*) was used to assess decoding performance.

Non-face identity decoding. PS's impairments for individuation of visual objects are restrained to faces (Rossion, 2018, 2022a). As a standard of comparison to the neural face-identification

decoders, we trained multiclass linear discriminant classifiers decoders to predict 8 within-category identities from *non-face* categories (i.e. objects, scenes, and animals). These classifiers were trained with identical parameters to the face-identity classifiers. Separate classifiers were trained for each category (e.g. a decoder for 8 identities within the animal category, another for 8 identities within the object category, another for 8 identities within the scene category; ~260 observations per session per participant). The time courses of these decoders' performance were averaged across all three categories and sessions. Recall was again used to assess decoding performance.

Comparison of brain representations with computational models

We compared our participants' brain RDMs to those from the vision (Fig. 6.3a) and caption-level description (Fig. 6.3b) models described in the previous section using Spearman correlations. Partial correlations were used where mentioned.

Reliability of brain representations across recording days

We computed the reliability of brain representations in a similar way as in Charest et al., 2014 (Charest et al., 2014a). For each participant, two RDMs were computed from two separate recording days, and compared using Spearman correlation in a time-resolved manner. Significance was assessed using permutation testing. Specifically, we created, for each participant and at each 4 ms step, a null distribution of 1000 brain-to-brain correlations using an RDM in which rows and columns indices were randomly shuffled.

Group comparison and inferential statistics

All contrasts between PS and neurotypical controls were computed using Crawford-Howell modified t-tests for case-controls comparisons (Crawford & Garthwaite, 2012; Crawford & Howell, 1998). All time-resolved contrasts were computed from 0 to 1.1 s after image-onset. Spearman correlations were used for correlations within the control group, i.e. correlation analyses with behavioural performance.

Permutation testing was used throughout the paper to assess significance of EEG-RDM to computational model latencies. We created, for each participant and at each 4 ms step, a null distribution of 1000 brain-model correlations using model RDMs in which rows and columns indices were randomly shuffled (Kriegeskorte, Mur, & Bandettini, 2008).

Permutation testing was also used to assess significance of identity decoding latencies. We created, for each participant, session, and at each 4 ms step, a null distribution of 500 decoding performances using identity labels which indices were randomly shuffled. The average of these distributions across sessions and participants were used to assess significance of the time-resolved decoding performance shown in **figure 6.3a**.

Results

One-back task

Accuracies did not differ between aged-matched and young controls sub-groups either for face stimuli ($t(16) = -0.3099$, $p = .761$; $t(16) = -0.9607$, $p = .3510$) or non-face stimuli ($t(16) = 1.2925$, $p = .215$; $t(16) = -.09704$, $p = .3463$). Therefore, their data have been aggregated into a single neurotypical control group. PS differed significantly from controls on a face vs. non-face performance score, computed as the first PCA component of face vs. non-face accuracies and response times ($t(17) = -7.7157$, $p = 1.6053 \times 10^{-6}$).

Cambridge Face Memory test long-form (CFMT+)

Within the control participants, CFMT+ scores did not differ between aged-matched and young controls sub-groups (t(16) = -0.8058, p = .4322). Their data have been aggregated into a single control group. PS significantly differed from controls on this standard face identification ability score (t(17)=-2.7623, p = 0.0133).

To assess the face-specific performance of PS and controls in a single individual score across all behavioural measures, we combined performance in the one-back task (accuracies and response times of face and non-face trials) and face-memory performance (CFMT+) using Principal Component Analysis (PCA). Specifically, face-specific performance in the one-back tasks was computed as a face vs. non-face performance score ([face - nonface]./[face + nonface]) separately for accuracy and RTs, for each participant. We used PCA to extract projections explaining variance across these two variables as well as the CFMT+ (Calder et al., 2001; Calder & Young, 2005). The first component, which explained 65.76% of the variance in performance across participants, is henceforth referred to as the face-specific performance score. PS significantly differed from neurotypical controls on this score (t(17) = -7.1966, p = 1.0691e-06; see **figure 6.2a**), indicating typical face-specific behavioural deficits in this patient.

Reliability of neural dynamics measured with RSA

Measuring brain dynamics of brain-lesioned patients can be arduous using raw electrophysiological topographies (Alonso Prieto et al., 2011). We assessed whether we could measure reliable brain representations of brain-lesioned PS by computing inter-session reliability of brain Representational Dissimilarity Matrices (RDMs). We computed the correlation between EEG RDMs computed from recording day 1 and recording day 2 at every 4 ms steps. The timecourse of these correlations, shown in figure 6.2a, indicates significant (ps<.05, permutation testing) reliability of brain representations of both neurotypicals and PS across most time points after image onset, peaking in the N170 window (160 ms and 180 ms for controls and PS,

respectively; r_{peak_ctrls} = .4316; r_{peak_PS} = .2638). PS showed surprisingly high SNR across sessions, her correlation time course being indistinguishable from the neurotypicals' reliability scores across time points (i.e. no significant contrasts).

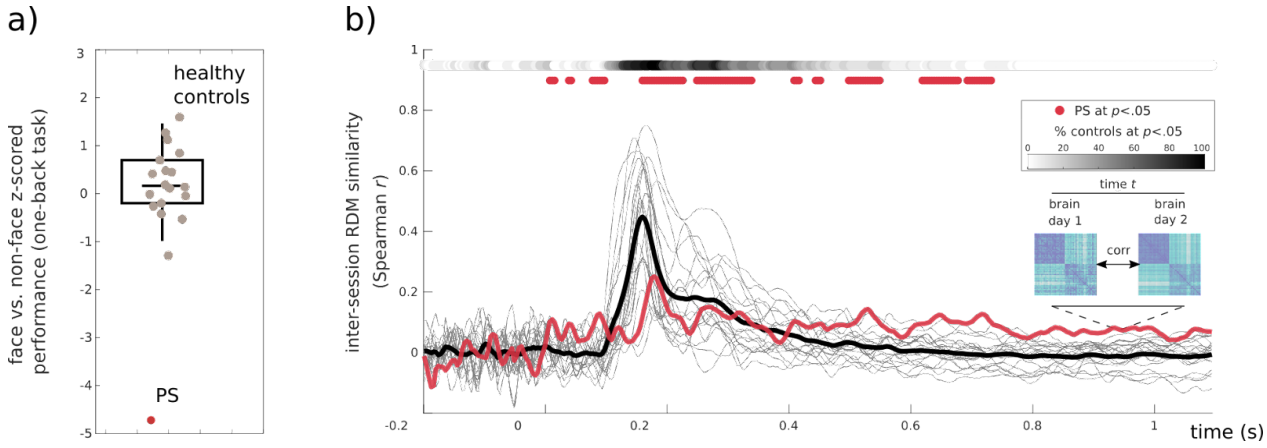


Figure 6.2. Behavioural performance and reliability of brain representations across time. (a) A behavioural face-specific performance score was computed for all participants and constrasted between PS and controls, indicating strong and archetypical face-sensitive deficits (p<.0001). **(b)** Brain Representational Dissimilarity Matrices (RDMs) computed on different recording days were cross-correlated within participants for patient PS (red line) and neurotypical participants (grey lines indicate individuals participants, black line indicate control-averaged). PS brain RDMs showed significant inter-session reliability across most time windows (p<.05, permutations). For comparison, the percentage of control participants with significant cross-session correlations is shown on top for every time point, with darker points indicating higher percentage of neurotypicals with reliable brain RDMs. PS had similar reliability coefficients compared to controls.

Impaired neural decoding of face-identity

We first asked whether we could capture PS's face-identification deficits at the neural level (Gao et al., 2019; Liu-Shuang et al., 2016a). For each participant, we decoded face-identity from brain activity by training a multiclass linear discriminant classifier from whole-brain EEG patterns at each 4 ms steps from face onset (see **figure 6.3a**). In neurotypical controls, this resulted in weak

but above chance decoding of face-identity (from 138 ms after face onset, peaking at around 180 ms; *ps*<.05, permutations; see **figure 6.3a**), as can be expected from this difficult classification task (Dobs et al., 2019; Kriegeskorte et al., 2007; Liu-Shuang et al., 2016b; Muukkonen et al., 2020; Tsantani et al., 2021). PS's face-identity decoding was continuously at chance level (no significant time points after face onset; *ps*>.05, permutations). Time-resolved contrasts with controls showed reduced neural identity decoding in PS around 180 ms (Howell-Crawford t-tests; *ps*<.05). Neural decoding of *non-face* identities, on the other hand, appeared entirely spared in PS. Non-face identity classifiers resulted in above chance performance in both controls *and* PS. Furthermore, PS's neural individuation for non-face objects was within the normal range of controls across all time points after face onset (no significant contrasts PS < controls; see **figure 6.3b**). Peak individual neural decoding evidence for face identities (**figure 6.3c)** and non-face identities (**figure 6.3d**) across participants showed PS positioned at the low end of the spectrum of the neural face individuation scores ($t(18) = -2.48$, $p < .05$), but showed that she had a typical score on the neural *non-face* individuation across our sample ($t(18) = -0.73$, $p = .47$).

We further tested the behavioural relevance of this decoded face identity evidence. We correlated the magnitude of neural decoding across participants (i.e. the neural individuation score) with behavioural face-identification abilities, as measured by the CFMT+. Notably, this was done *with individuals from the control group only.* This showed that, indeed, neural face-individuation score correlated positively with face recognition abilities across neurotypicals (peak-r = 0.63, peak-r_{time} = 330 ms; *ps*<.05, ~174-702 ms; **figure 6.3c**; *p*<.05, permutations). Identical brain-behaviour correlation analyses using *non-face* individuation scores resulted in much reduced correlations with the CFMT+, with significant positive correlations only appearing in a short window around 100 ms after non-face objects onset (**figure 3d;** *p*<.05, permutations). Both the contrasts between PS and controls and these interindividual results were replicated using the brain RDMs of our participants and an identity model RDM (Dobs et al., 2019; see **Supplementary material**).

Thus, we were able to capture the poor neural identity representations of PS using

multivariate EEG signal, showing her reduced neural distance between face identities around 180 ms. This impairment appeared specific to face-individuation, and the magnitude of this representational distance, particularly around 330 ms, was predictive of neurotypical face-processing abilities.

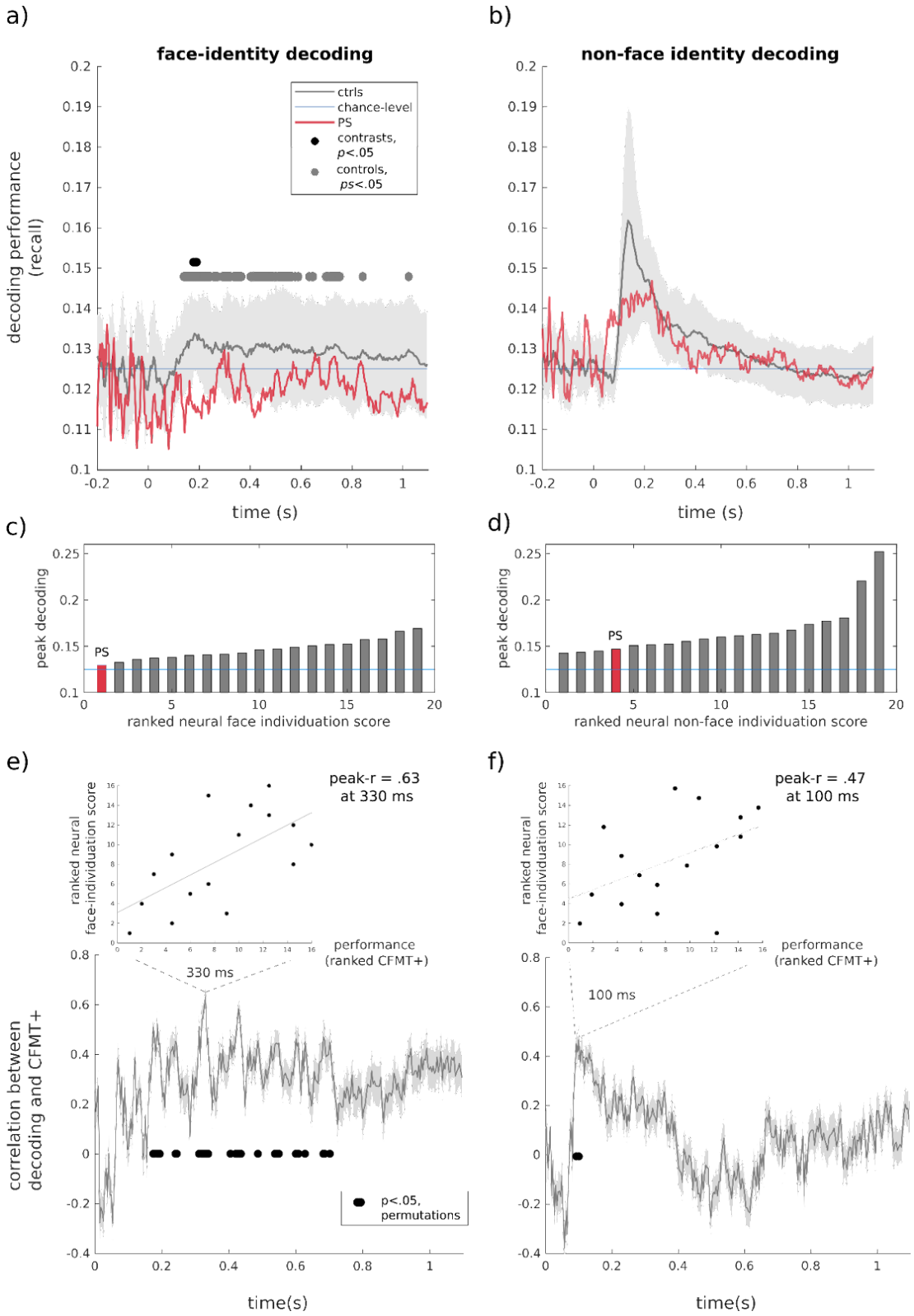

a)
face-identity decoding
ctrls
chance-level
PS
contrasts, p<.05
controls, ps<.05
decoding performance (recall)
time (s)
b)
non-face identity decoding
time (s)
c)
peak decoding
PS
ranked neural face individuation score
d)
peak decoding
PS
ranked neural non-face individuation score
e)
peak-r = .63 at 330 ms
ranked neural face-individuation score
performance (ranked CFMT+)
330 ms
correlation between decoding and CFMT+
p<.05, permutations
time(s)
f)
peak-r = .47 at 100 ms
ranked neural face-individuation score
performance (ranked CFMT+)
100 ms
time(s)

Figure 6.3. Neural decoding of face and non-face identities and association with individual face recognition abilities. Multivariate classifiers were trained to predict 8 face identities **(a)** as well as 8 identities from non-face categories **(b)** from time-resolved EEG patterns. **a)** Neural face-identity decoding performance (recall, 5-fold, 5-repetition cross-validation; Treder, 2020) is shown for PS (red line) and the control group (black line, representing control-average decoding time course; grey dots show significant identity decoding at $p<.05$, permutations). PS had overall below chance decoding for the face-identity decoding across time and showed significantly lower identity decoding compared to controls around 180ms after face onset (black dots indicate significant contrasts at $p<.05$). Shaded error bar represents standard deviation within the control group. **b)** PS showed typical neural decoding of non-face identities across all time points, with no significant contrasts with controls. Peak individual neural decoding evidence for face identities **(c)** and non-face identities **(d)** have been ranked across participants, and demonstrated similar effects with PS at the lower end of the spectrum of the neural face individuation scores ($t(18) =-2.48$, $p <.05$), but with a typical score on the neural non-face individuation ($t(18) = -0.73$, $p=.47$). **e)** Spearman correlation between individual neural decoding evidence and face recognition abilities (i.e. the CFMT+ scores across neurotypical participants only) was computed at each 4 ms step. Correlation time course and shaded error bar were computed using a jackknife procedure. Significant brain-behaviour correlations (permutation testing at $p<.05$) were found in several time points for the face-identity decoding (between 174 ms and 702 ms, peaking at 330 ms). **f)** In contrast, brain-behaviour correlation with non-face identity decoding and CFMT+ scores were restrained to a small early window peaking at 100 ms (92-100 ms). Each panel shows a scatter plot of CFMT+ and neural decoding computed at peak correlation with corresponding least-square regression line.

Similarity with visual and semantic computational models

Next, we characterised the specific neural computations underlying these deficits. We assessed visual brain computations in PS and neurotypical controls by comparing their brain RDMs to those of convolutional neural networks (CNNs) trained to categorise objects (Güçlü & van Gerven, 2015; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014). The visual model RDMs were produced for all 8 layers of the CNN (see **supplementary Fig 6.2**). Controls showed significant brain-RDM correlations with the sixth layer of the visual CNN around 98 ms (CNN layer 6, $p<.05$, permutations), continuing as late as around 500 ms after image onset (see **Figure 6.4a;** CNN layer 3 : 59 - 746 ms; CNN layer 1 : 100 - 242 ms). PS showed significant correlations with the CNN layer 6 in a more restrained window between 153 and 204 ms (layer 1: ns; layer 3: 211 - 215 ms). Direct contrasts of PS' correlation time courses with those of controls indicated reduced correlations with RDMs across the visual CNN (i.e., at layers 1, 3, 6 & 7;

Figure 6.4a, $p<.05$). These contrasts peaked at layer 6, which represent a higher proportion of high-level visual features (e.g., whole objects and object parts, Güçlü & van Gerven, 2015; Long et al., 2018).

To reveal whether even higher-level semantic computations (Barton et al., 2009; Schweinberger & Neumann, 2016) could be affected in the brain of PS, we used a deep averaging network (Google Universal Sentence Encoder, GUSE; Cer et al., 2018) to transform human-derived captions of our stimuli (e.g. " a city seen from the other side of the forest ") into embeddings (points in a caption space). Then, we compared the RDMs computed from this semantic model to the brain RDMs of PS and control participants. Controls showed significant correlations with semantic computations from around 98 ms, continuing as late as 981 ms after image onset, and peaking at 164 ms (peak-r_{ctrls} = .45; $p<.05$, permutations, **Figure 6.4b**). Again, PS showed significant correlations with this model in a more restrained window from around 176 to 200 ms, and peaking ~20 ms later than controls at 188 ms (peak-r_{PS} = .173; $p<.05$, permutations, **Figure 6.4a**). Direct contrasts confirmed the reduced correlations with these semantic computations in the brain of PS compared to controls (**Figure 6.4b**, $p<.05$). Note that this comparison, as well as the one with the visual model, excluded the information shared between the semantic and visual models. Similar results were observed when comparing brains and the visual and semantic model without removing the shared information between the brain and the visual CNN (see **supplementary Fig 6.3 & Supplementary Fig 6.4**).

A summary of significant contrasts with all computational models (**figure 6.4c**) indicates that PS's brain processing stream exhibited impairments peaking in higher-level visual (CNN layer 6) and semantic (caption-level) representations.

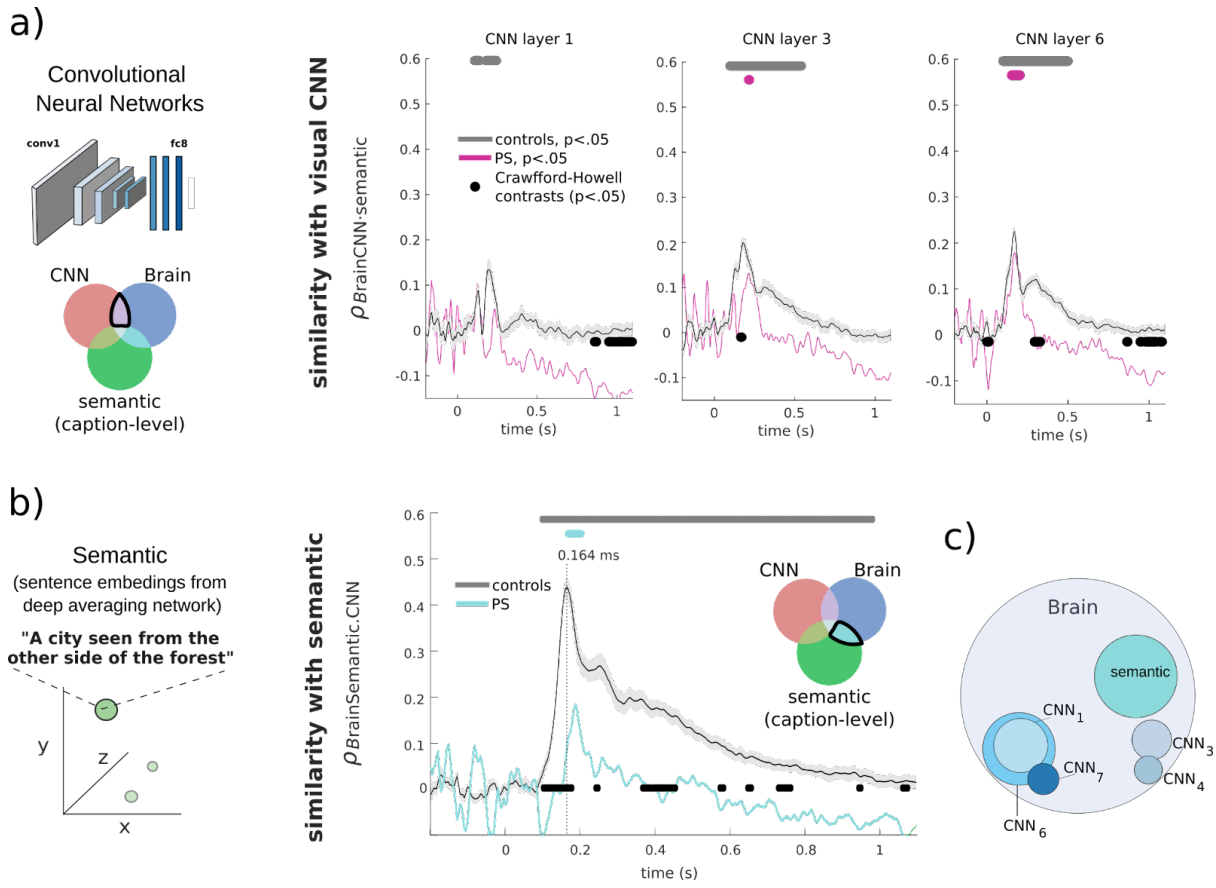


Figure 6.4. Comparison of brain representations with those of artificial neural networks of visual and semantic processing. a) Partial correlation between brain RDMs and AlexNet RDMs (removing shared correlation between brain and semantic model) is shown for PS (coloured curve) and controls (grey curve). Each column shows different layer RDMs in ascending order from left to right. We found overall lower similarity of visual computations within the brain of PS compared to controls (black dots indicate significant contrasts, Howell-Crawford modified t-tests, $p<.05$), with differences peaking in higher-level CNN layer 6. Similar results were observed when comparing brains and CNN models without removing the shared information between brains and the semantic (caption-level) model (see supplementary figure 2). **b**) Partial correlation with RDMs of the semantic model (excluding shared information between brain and AlexNet) was significantly lower in the brain of PS compared to controls (cyan curve; black dots indicates significant contrasts, $p<.05$). The shaded areas of all curves represent the standard error for the controls. **c**) Proportion of all time points after EEG onset showing significant contrasts (PS < controls, $p<.05$, uncorrected) is shown on a Venn diagram as a summary of PS's impaired brain computations for visual computations (blue colours, where deeper blue represents deeper CNN layers) and semantic computations (cyan). The total surface area for the "brain" section represents all possible time points after onset (0-1100 ms in 4 ms steps), while the surface of each categories represents a portion of these time points. Overlap indicates significance over the same time points.

Relationship with electrophysiological brain components

We specified the time at which visual and semantic brain computations differed in PS by producing brain-to-model similarities in time windows corresponding to well-known Event-Related Potential (ERP) components indexing early, mid, and late neural processing, i.e. the P100 (Luck et al., 1990), N170 (Bentin et al., 1996) and N400 components (Kutas & Federmeier, 2000), respectively (**figure 6.5a**). Within neurotypicals, we found EEG representations peaking in similarity with the visual CNN at mid-layers (fourth and fifth; Jiahui et al., 2022) around mid-level temporal windows. Similarity with semantic computations also peaked around mid-latencies. EEG-correlations with the semantic model, however, surpassed those with the visual CNN at mid and late latencies (paired t-tests comparing semantic vs. visual correlation, *ps* <.05).

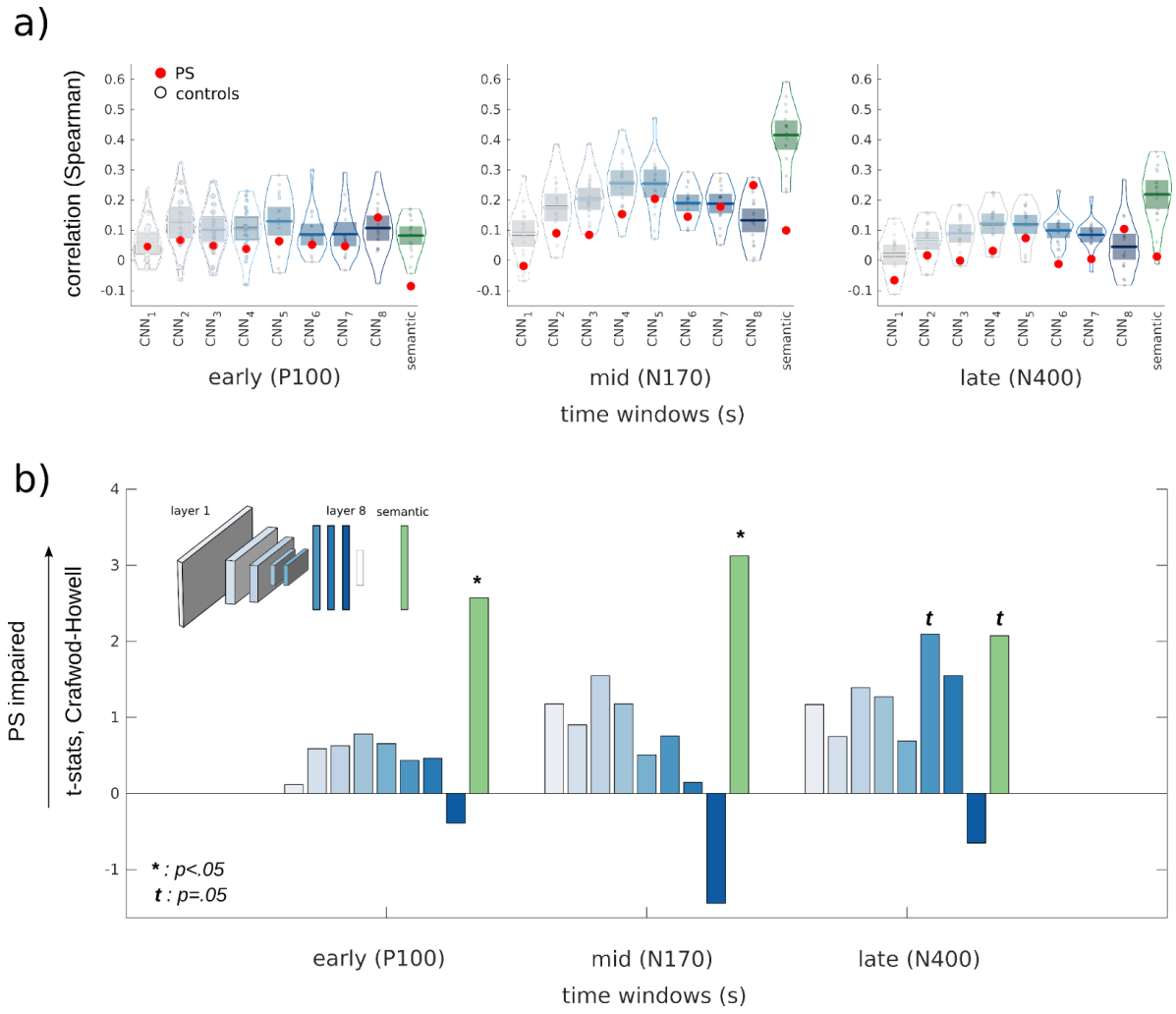


Figure 6.5. Comparison of visual and semantic computations along early, mid and late brain processing steps. (a) Correlations of brain RDMs to visual model (blue plots, with deeper blue corresponding to deeper CNN layers) and semantic model (green plot) was assessed in PS and controls for brain RDMs computed in temporal windows corresponding to the P100 (80-130 ms), N170 (130-230 ms) and N400 (250-500 ms) ERP components. **(b)** Crawford-Howell t-statistics comparing controls vs. PS on these correlations are shown for the visual CNN (blue bars, with deeper blue corresponding to deeper CNN layers) and the semantic model (green bars; significant comparisons are indicated with an asterisk, marginally significant comparisons with the letter *t*).

In contrast to this tendency, PS showed generally lower similarity to semantic computations across all time windows. Direct contrasts of brain-model similarity between controls and PS are shown in **Figure 6.5b**, with higher values indicating stronger impairments in

PS in visual/semantic computations. We observed reduced semantic brain computations for early and mid-processing windows around the P100 and N170 components (Bentin et al., 1996; Clark et al., 1994), respectively, as well as a marginally significant effect in late windows around the N400 component (p = .0536; Kutas & Federmeier, 2000). Contrasts of similarity with the visual CNN layers were overall less pronounced after averaging RDMs in these ERP windows. Similarity to the CNN layer 6 showed the only marginally significant difference between PS and controls around the late processing window corresponding to the N400 (p = .0512). More fine-grained temporal contrasts using time windows of 60 ms steps (see **supplementary figure 6.5**) confirmed these results.

As stated earlier, a predominant assumption in cognitive-neuroscience is that early brain signal refers to low-level computations while later brain signal refers to higher-level computations (e.g. Wiese et al., 2019). To assess whether such progression was present in the neural time course of our participants, we computed all pairwise correlations between brain RDMs, CNN-layers RDMs, and semantic RDMs, resulting in a second-level representational similarity matrix (RSM) of 12 x 12 spearman-correlation values (**figure 6.6a**). Comparing all RDMs in this manner is similar to the analyses described in the previous section, but allows two additional insights. First, it reveals how similar brain representations computed from different temporal windows are with one another (King & Dehaene, 2014). In our setting, this showed that neural representations during the N170 are more similar to those during the N400 compared to earlier (P100) computations within controls. Second, this analysis reveals how similar representations of computational models are along their putative hierarchical computational stream (i.e. layers 1-8 & semantic processes; (Güçlü & van Gerven, 2015). This showed, reassuringly, that the representations of the CNN used here are also generally more similar in layers closer apart within the CNN's architecture (e.g. layer 6 and 7 are more similar to one another than layer 1 and 7). Finally, both of these levels of analysis and their interaction can be summarised by computing the optimal two-dimensional solution of this multidimensional space

with multidimensional scaling (MDS). This visual summary is shown in **figure 6.6b**. Because it produces 2D coordinates in a (representational) space as it unfolds across time, producing a path or progression of brain representations, it will be referred to as representational trajectory. In both PS and controls, this analysis confirmed a trajectory from early brain representations, relatively more similar to early layers of the visual CNN, to mid and late brain representations, relatively more similar to the deeper layer of the CNN and semantic model. Some differences, however, are noticeable. First, these second-level RSMs indicated the reduced magnitude of similarity to the models in PS outlined in the previous section, specifically to those of higher-level semantic computations. More importantly, these representational trajectories revealed that PS' late neural code is both i) more similar to her early (brain) representations around the P100, and ii) more similar to low-level (visual) computations of the CNN. These results suggest relatively less change in neural computations from early to late stages of PS's visual stream (Cichy & Oliva, 2020).

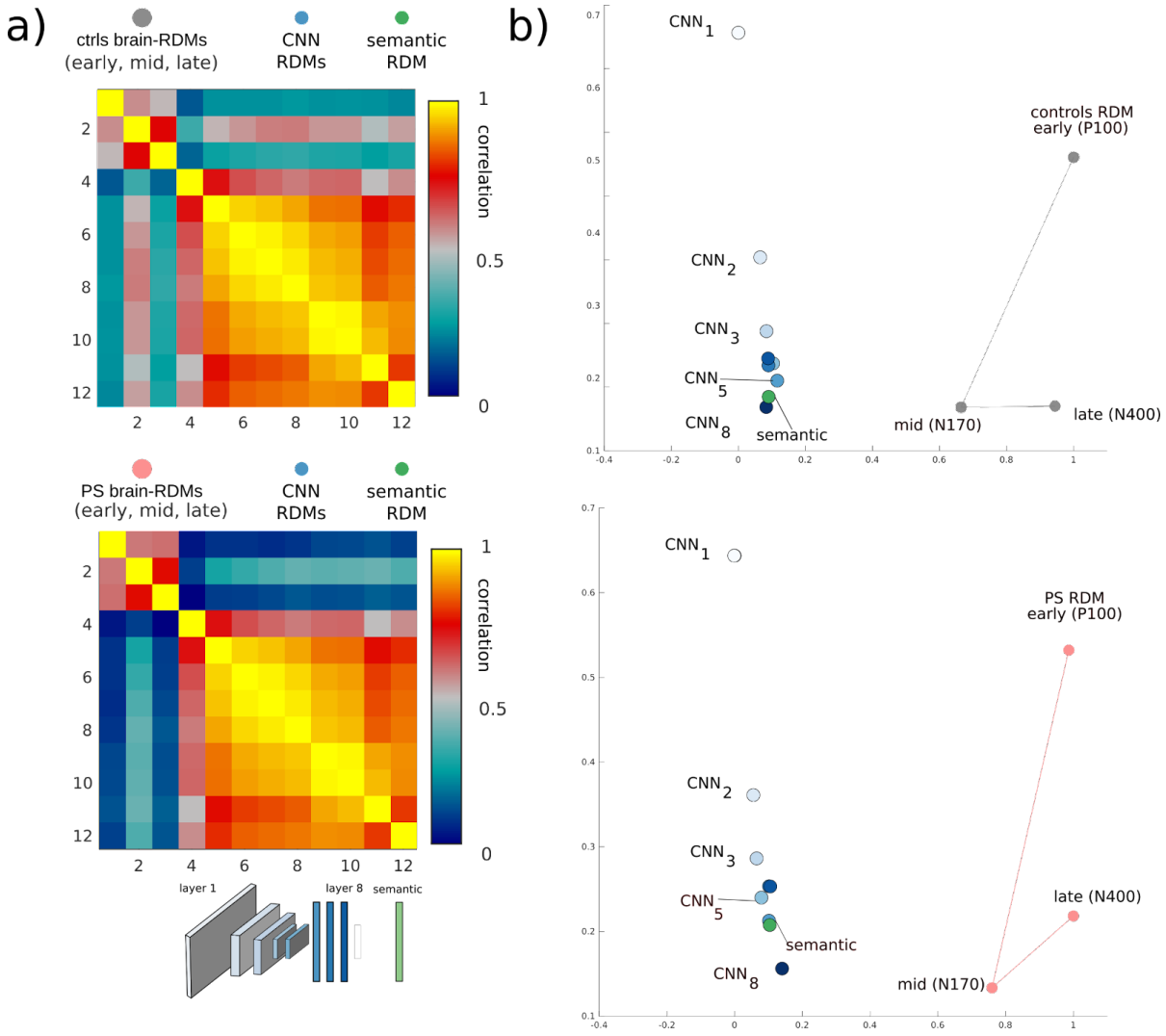


Figure 6.6. Representational trajectories of brain and computation models. (a) All pairwise correlations across brain and computational models' RDMs resulted in second-order similarity matrices (RSMs) for controls (upper panel) and PS (lower panel). **(b)** A summary of group RSMs was obtained by computing the 2D solution of these RSMs using MDS. Points corresponding to brain representations of controls (upper panel) and PS (lower panel) were linked with grey and red lines, respectively.

Discussion

Here, we characterised the neural computations of a well-studied single-case of acquired pure

prosopagnosia, patient PS, using whole brain multivariate signal and computational models of vision and caption-level semantics. We were able to capture the impaired functional neural signatures of face individuation–as well as the absence of non-face individuation deficits–in the brain of PS. Most importantly, we show that PS' inability to identify faces is associated with reduced high-level visual computations, including a significant reduction of even higher-level semantic brain computations. This reduction in semantic computations was present as soon as the P100, peaked around the well-known N170 component, and persisted later around the N400 window. Analyses of brain dynamics revealed that the neural code of patient PS appears more stable across time compared to healthy controls, with fewer changes from early to late brain representations.

Impairment for face individuation is the archetypal deficit of patient PS and prosopagnosia in general (Bodamer, 1947), but has been arduous to reveal solely from brain activity (Anzellotti et al. 2014; Alonso Prieto et al. 2011). Using a data-driven whole-brain approach, we extend recent work having been able to reveal such deficits from the neural responses of PS with fast periodic visual stimulation (FPVS; Liu-Shuang et al., 2016b) by describing the temporal unfolding of this deficit under normal passive viewing conditions. We found reduced neural evidence for face identity around the N170 window (Bentin et al., 1996) in PS, and, conjointly, revealed normal *non-face* individuation in this patient (Dricot et al., 2008; Liu-Shuang et al., 2016b). We further discovered that variations in typical face recognition abilities, as indexed by gold-standard behavioural tests in control participants (Russell et al., 2009), is associated with this neural decoding of face-identities from around the N170 onward, peaking around 350 ms in the N400 window (Johnston et al., 2016; Schweinberger & Neumann, 2016; Wiese et al., 2019b). Again, individual neural decoding of non-face individuation did not correlate with neurotypical face-recognition abilities in these time windows but was rather associated with early neural activity around 100 ms after face onset.

Crucially, by associating electrophysiological signal and computational models, we found that the underlying neural computations of PS differed most with respect to the higher-level

visual and semantic computations of deep neural networks models (DNNs). The late layers of visual DNNs have been previously linked to processing in human infero-temporal cortex (hIT; (Güçlü & van Gerven, 2015; Jiahui et al., 2022.; Khaligh-Razavi & Kriegeskorte, 2014), peaking in the FFA (Khaligh-Razavi & Kriegeskorte, 2014), and functionally to higher-level visual feature representations such as parts of objects, whole objects and viewpoint invariant representations (Güçlü & van Gerven, 2015). These observations are consistent with the impaired whole face (Ramon et al., 2016) and feature representations (Caldara et al., 2005; Fiset et al., 2017) previously described in patient PS. Associations between brain activity and higher-level semantic computations have only been attempted much more recently (Dwivedi et al., 2021; Faghel-Soubeyrand et al., 2022b; Popham et al., 2021). The computations revealed here are arguably closer to visuo-semantic representations, which have been shown to be represented both within the visual ventral stream (including the FFA and OFA; Doerig, Kietzmann, et al., 2022; Popham et al., 2021) and outside (Doerig et al. 2022). Using a model of caption-level semantics (Cer et al., 2018; Faghel-Soubeyrand et al., 2022; Doerig et al. 2022), we show that PS's representational geometry displays significant — but greatly reduced — correlation with these semantic computations compared to controls. These findings demonstrate a clear link between semantic brain computations and important changes in face recognition ability (Bruce & Young, 1986; Duchaine & Yovel, 2015).

Together with similar computational characterisation of brain representations in individuals with "super-recognition" of faces (Faghel-Soubeyrand et al., 2022; Russell et al., 2009), our results suggest *some* gradient of neural computations from the low-end to high-end of face-recognition abilities. Indeed, super-recognisers in (Faghel-Soubeyrand et al., 2022) showed *enhanced* similarity with mid-level visual and semantic computations around the N170 and P600 windows, respectively. Here, while PS showed reduced similarity with visual and semantic computations, obvious differences are also apparent. PS's neural computations, particularly semantic computations, appear to be impacted much sooner and on a larger extent than those of super-recognisers. Whereas semantic brain computations were enhanced around the P600 in

super-recognisers, here PS displayed reduced similarity as early as the P100, continuing throughout the N170 and N400 windows. In fact, PS showed no significant correlation with either the visual or semantic-level computations at any point during late processing windows (i.e. after ~300 ms). These observations suggest that PS's deficits in neural computations start relatively early along the typical processing stream. Some differences, however, are to be expected given the important lesions to the face processing network (i.e. including the right-OFA and left-FFA) of this patient. Indeed, the OFA is not only involved in the extraction of featural information about faces (Pitcher et al., 2007), but has also been causally linked to higher-level identity (Ambrus et al., 2017; Jonas et al., 2014) and semantic information (Eick et al., 2020) processing. Both these associations and the fact that the Anterior Temporal Lobe (ATL) — an important hub for the processing of semantic information — is preserved in PS (Rossion 2022; Gao et al. 2019) points to a similar involvement of the OFA/FFA complex in the semantic neural computations revealed here. Future work using brain imaging with higher spatial resolution will be required to address these questions.

While the present study offers insights on the neural dynamics of prosopagnosia, it is also limited in a number of ways. For example, even though we recorded a sizable data set of high-density EEG comprising numerous trials and stimulus conditions, signal-to-noise ratio was still small when probing specific neural processes. Characterising face identity representations across time is a notoriously difficult task (Anzellotti et al., 2014; Dobs et al., 2019) that could have benefited from recordings of even higher density imaging (M/EEG in Dobs et al., 2019; Vida et al., 2017; fMRI in Nestor et al., 2011), as well as from the use of experimental tasks recruiting more directly this process (Nemrodov et al., 2016). Another related and potentially fruitful approach to reveal brain computations and representational geometry in prosopagnosia would be to produce EEG-fMRI "fusion" (Cichy et al., 2014; Cichy & Oliva, 2020). Indeed, the spatio-temporal description procured by linking EEG *and* fMRI data with RSA would be ideal to test hypotheses regarding changes in neural code such as the one hinted in our findings. Employing this technique on PS's EEG and additional fMRI data could reveal the putative

hurdles in the transformation of neural code across the different lesions of PS's, and thereby attribute causal links between OFA/FFA and impairments in brain computations. Finally, comparison of brain representations with computational models in general also has certain limits (Doerig, Sommers, et al., 2022). Recent advances in developing more interpretable (Schyns et al., 2022; Soulos & Isik, 2020) and ecological models (Kietzmann et al., 2019; Mehrer et al., 2021) of human cognition will be helpful to address them.

Notwithsanding these limitations, this work offers, to our knowledge, the first description of the fine-grained temporal processes combined with a state-of-the-art computational characterisation of the brain representations in prosopagnosia. Understanding the very nature of defective perceptual representations offers new routes for patient rehabilitation. Indeed, we believe that the recent technological advances that permitted us to reveal these findings offer new promising ways to understand, and could perhaps even help diagnose the fine-grained deficits in perception and cognition in diverse clinical populations.

Acknowledgements

We sincerely thank PS for her precious contribution and participation in this study.

Funding

Funding for this project was supported by an ERC Starting Grant [ERC-StG-759432] to I.C, an ERSC-IAA grant to J.W., I.C. and S.F.S., by a Swiss National Science Foundation grant (10001C_201145) to A.-R.R. and R.C., and by a NSERC and IVADO graduate scholarships to S.F.S.

Author contributions

(CRediT standardised author statement)

S.F-S. : Conceptualisation, methodology, software, formal analysis, investigation, data curation, writing - original draft, visualisation, supervision, project administration, funding acquisition. **A-R.R.**: project administration, investigation, writing - editing and reviewing. **J.W.:** funding acquisition. **D.W. :** Investigation **R.C. :** Resources, project administration, writing - editing and reviewing. **F.G.** supervision, writing - editing and reviewing. **I.C.**: Supervision, Methodology, resources, formal analysis, writing - editing and reviewing, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Data availability

Data are available from the corresponding authors upon request.

Code availability

The MATLAB and Python codes used in this study will be available upon request.

Supplementary material

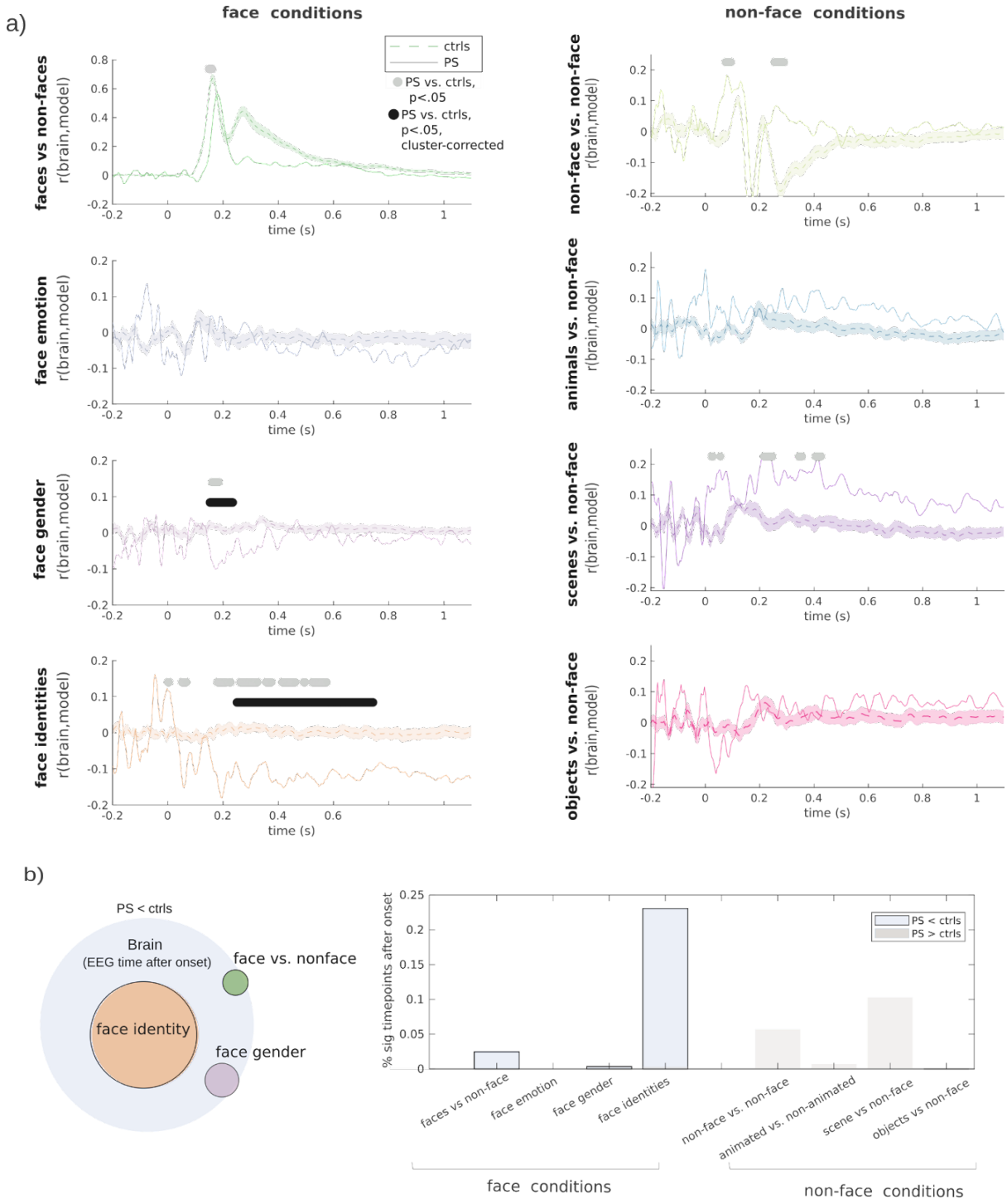
Neural dynamics for key categorical distinctions

We derived neural distances between face-identities by correlating the brain RDMs of our participants with a categorical face identity model RDM (where ones indicated multivariate distances between different face identities, and zeros indicated distance between same face identities). PS showed overall lower fit with this identity model from approximately 172 ms to 746 ms (see supplementary figure 1, $p<.05$, cluster-corrected), indicating extended impairment in neural processing of face identities (Johnston et al., 2016; Wiese et al., 2019a; Yan & Rossion, 2020). Notably, this reduced distances between face-identity representations cannot be solely attribute to signal-to-noise differences in the brain of the patient; PS RDMs showed on par correlations with other models of categorisation (e.g. scenes vs. non-faces, objects vs. non-faces; see supplementary figure 1) compared to typical participants.

We also tested whether these impaired face-identity representations in PS could generalise to neural processes of individuals with typical abilities, i.e. a critical prediction in the quantitative (vs. qualitative) accounts of individual differences in face-recognition ability (Barton & Corrow, 2016; Bobak et al., 2017; Hendel et al., 2019; Maguire et al., 2003; Price & Friston, 2002; Rosenthal et al., 2017; Vogel et al., 2005; Zadelaar et al., 2019). We correlated the individual neural individuation score across time with face recognition ability within the control participants only. This showed that neural distance between face identities correlated positively with face specific performance scores across neurotypicals (peak-r = 0.5305, peak-r_{time} = 418 ms; $ps<.05$ from 400-434 ms). Similar results were obtained with face vs. non-face accuracy in the one-back task, or the CFMT+ taken separately.

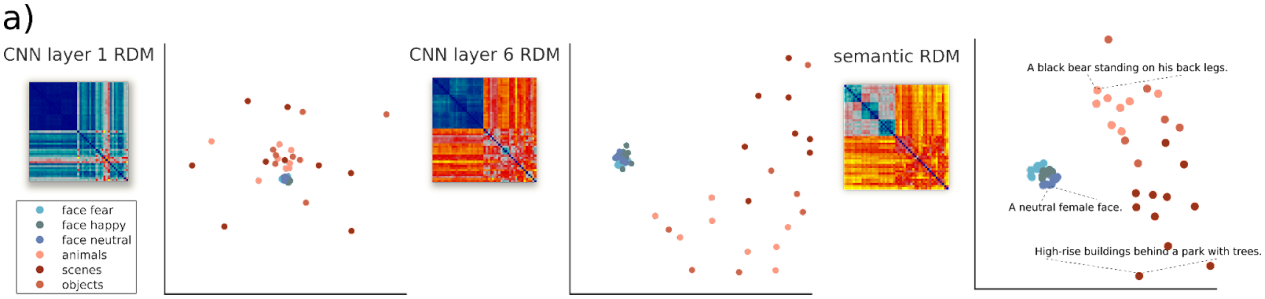
Previous studies have shown that prosopagnosic patient PS displays somewhat typical face-selective activation (i.e. typically higher face vs. non-face BOLD activity) throughout most

of the cortical face network (Gao et al., 2019). We asked whether fine-grained temporal multivariate analyses could reveal more subtle differences in the neural representations of PS. On top of face-identity, our various experimental conditions offered a mean to assess the putative typicality of PS's neural response for important categorical distinctions (e.g. face-gender, face vs. non-face, scenes vs. non faces, etc.). We derived RDM correlations with these categorical models in an identical way than with the face-identity model. For example, face-gender information was assessed as the fit between brain RDMs and a face-gender binary model (where ones indicated multivariate neural distance between faces of different sex; see (Dobs et al., 2019). As can be assessed in supplementary figure 1, while PS showed impaired neural fit across face models (particularly with face-identity, left column), she showed overall typical fits with the non-face models (right-column), thus replicating the decoding analyses shown in **figure 2**.

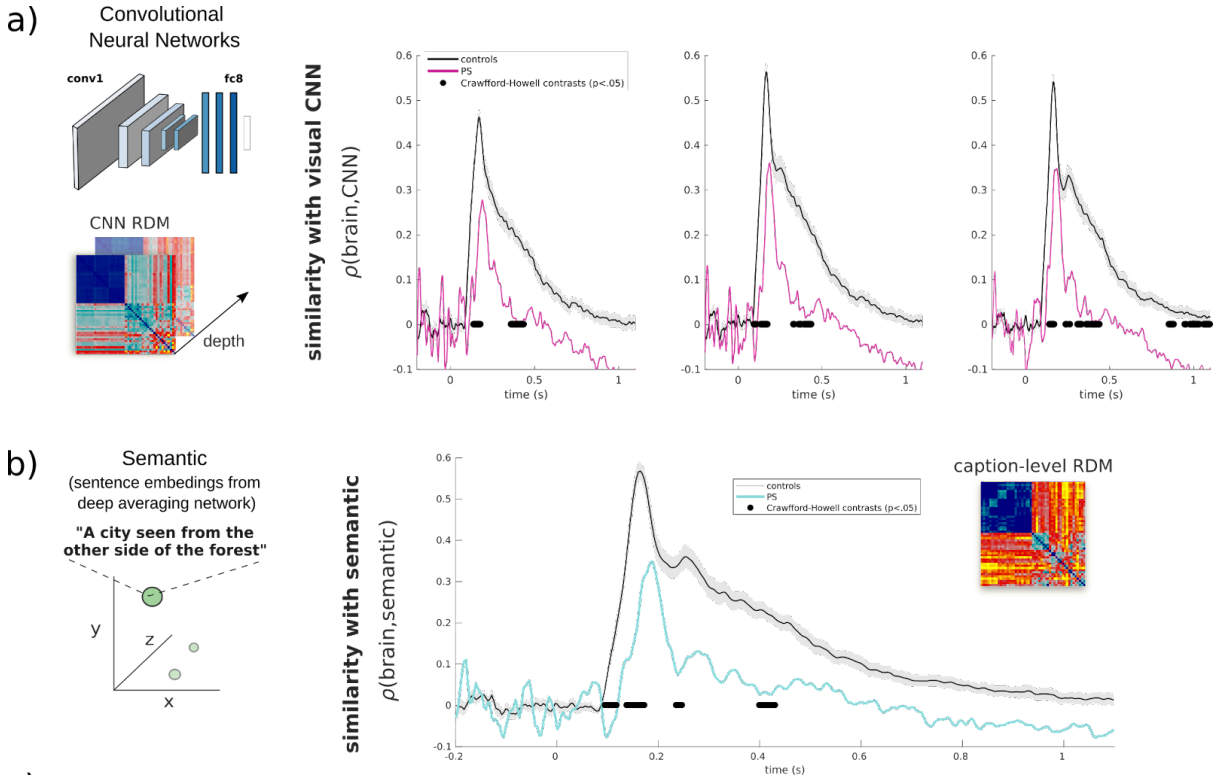


Supplementary figure 1. Time course of categorical distinctions compared between PS and neurotypicals. Brain RDMs for PS and controls were correlated at every 4 ms step with categorical models of face conditions (e.g., female vs. male faces in face gender condition) and non-face conditions

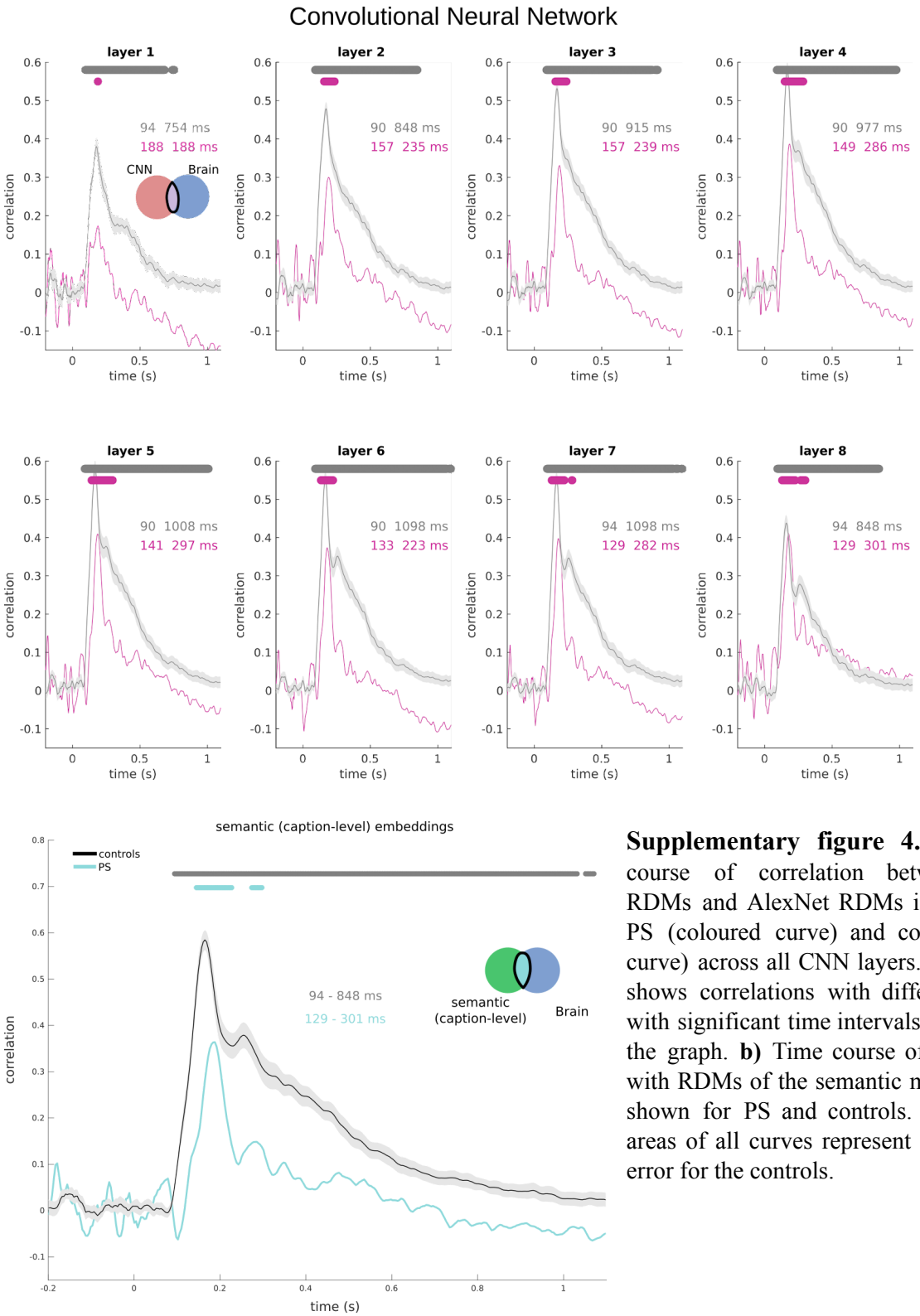
(e.g., scenes vs. non-faces). Comparison of these correlations between PS and controls showed significant contrasts (Howell-Crawford t-tests, $p<.05$, coloured points). These effects are summarized in panel b for significant contrasts where PS showed reduced fit compared to controls. PS showed reduced neural fit exclusively with face model conditions, while she showed better fit with scenes vs. non-face model conditions. Proportion of all time points after EEG onset showing significant contrasts (PS < controls, $p<.05$, uncorrected) is shown on a Venn diagram as a quantitative summary of PS's impaired brain computations for face identity (orange), face vs. non-face (green) and face-gender evidence (purple). The total surface area for the "brain" section represents all possible time points after onset (0-1100 ms in 4 ms steps), while the surface of each category represents a portion of these time points. Overlap indicates significance over the same time points.



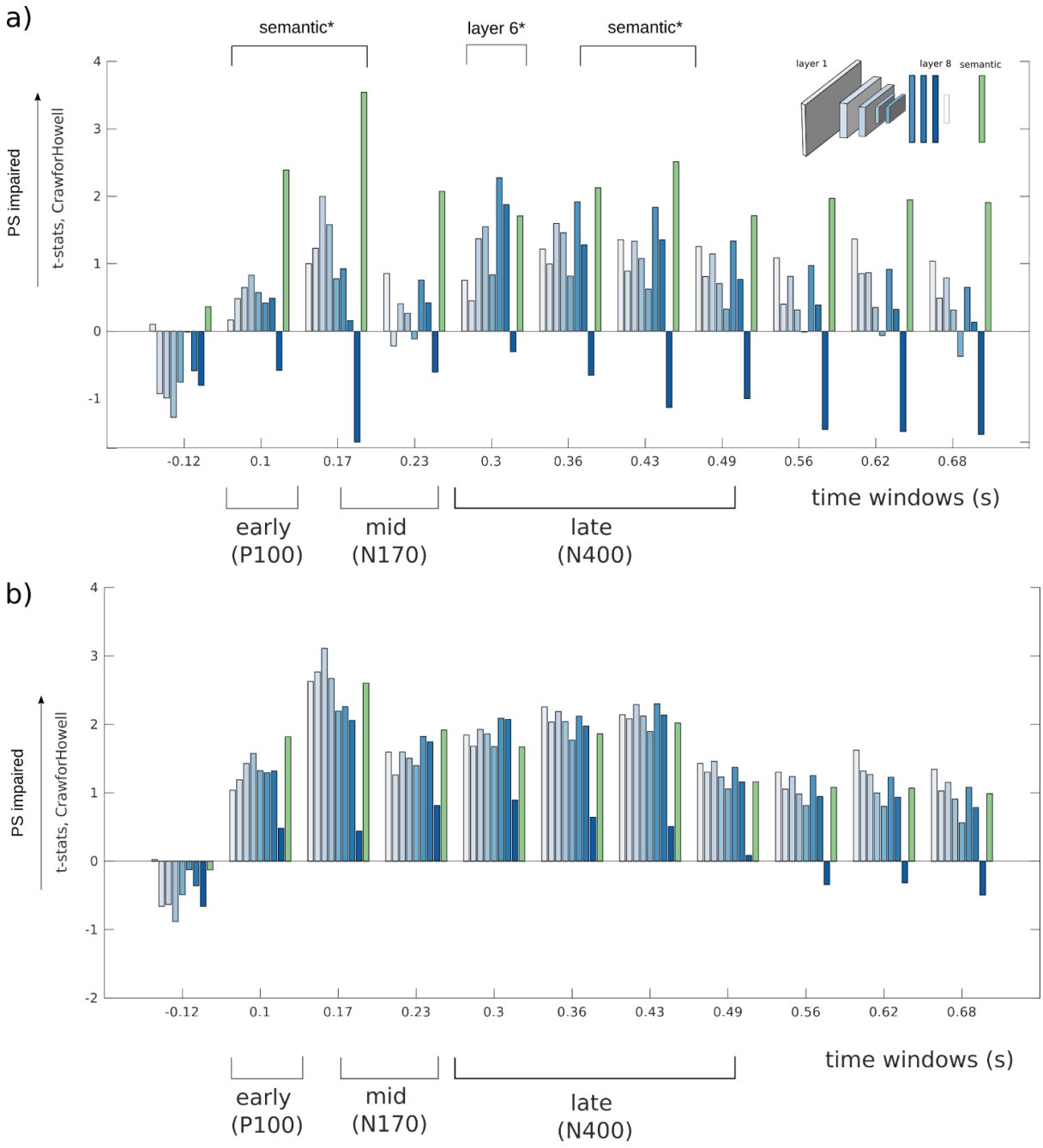
Supplementary figure 2. RDMs and 2D visualisation of example CNN layers and caption-level semantic model.



Supplementary figure 3. Time course of visual and semantic computations, unconstrained. a) Time course of correlation between brain RDMs and AlexNet RDMs is shown for PS (coloured curve) and controls (grey curve). Each column shows correlations with different layer RDMs 2, 4, and 6 in ascending order. We found overall lower similarity of visual computations within the brain of PS compared to controls (black dots indicates significant contrasts, Howell-Crawford t-tests, $p<.05$), indicating similar results to those observed when removing the shared information between CNN and the semantic (caption-level) model, in **figure 4**). Time course of correlation with RDMs of the semantic model was also significantly lower in the brain of PS compared to controls (cyan curve; black dots indicate significant contrasts, $p<.05$). The shaded areas of all curves represent the standard error for the controls.



Supplementary figure 4. **a**) Time course of correlation between brain RDMs and AlexNet RDMs is shown for PS (coloured curve) and controls (grey curve) across all CNN layers. Each graph shows correlations with different layers, with significant time intervals indicated in the graph. **b)** Time course of correlation with RDMs of the semantic model is also shown for PS and controls. The shaded areas of all curves represent the standard error for the controls.



Supplementary figure 5. Time course of categorical distinctions compared between PS and neurotypicals. Correlations of brain RDMs to visual model (blue plots, with deeper blue corresponding to deeper CNN layers) and semantic model (green plot) was assessed in PS and controls for brain RDMs computed at every 60 ms window. **(a)** **(b)** Crawford-Howell t-statistics comparing controls vs. PS on these correlations are shown for the visual CNN (blue bars, with deeper blue corresponding to deeper CNN layers) and the semantic model (green bars; significant comparisons are indicated with an asterisk). T-tests comparing partial correlations between PS and controls are shown in **(a)** and those comparing

simple correlations with the models are shown in **(b)**

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126. https://doi.org/10.1038/s41593-021-00962-x

Alonso Prieto, E., Caharel, S., Henson, R., & Rossion, B. (2011). Early (N170/M170) Face-Sensitivity Despite Right Lateral Occipital Brain Damage in Acquired Prosopagnosia. *Frontiers in Human Neuroscience*, *5*, 138. https://doi.org/10.3389/fnhum.2011.00138

Ambrus, G. G., Windel, F., Mike Burton, A., & Kovács, G. (2017). Causal evidence of the involvement of the right occipital face area in face-identity acquisition. In *NeuroImage* (Vol. 148, pp. 212–218). https://doi.org/10.1016/j.neuroimage.2017.01.043

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014). Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex* , *24*(8), 1988–1995. https://doi.org/10.1093/cercor/bht046

Avidan, G., Hasson, U., Malach, R., & Behrmann, M. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 2. Functional Neuroimaging Findings. In *Journal of Cognitive Neuroscience* (Vol. 17, Issue 7, pp. 1150–1167). https://doi.org/10.1162/0898929054475145

Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L. G., & Behrmann, M. (2014). Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral Cortex* , *24*(6), 1565–1578. https://doi.org/10.1093/cercor/bht007

Barton, J. J. S., & Corrow, S. L. (2016). The problem of being bad at faces. *Neuropsychologia*, *89*, 119–124. https://doi.org/10.1016/j.neuropsychologia.2016.06.008

Barton, J. J. S., Hanif, H., & Ashraf, S. (2009). Relating visual to verbal semantic knowledge: the evaluation of object recognition in prosopagnosia. *Brain: A Journal of Neurology*, *132*(Pt 12), 3456–3466. https://doi.org/10.1093/brain/awp252

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. https://doi.org/10.1162/jocn.1996.8.6.551

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*(1), 35–55. https://doi.org/10.1080/026432900380472

Benton, A. L., & Van Allen, M. W. (1972). Prosopagnosia and facial discrimination. *Journal of the Neurological Sciences*, *15*(2), 167–172. https://doi.org/10.1016/0022-510x(72)90004-4

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-Movement Strategies in Developmental Prosopagnosia and "Super" Face Recognition. In *Quarterly Journal of Experimental Psychology* (Vol. 70, Issue 2, pp. 201–217). https://doi.org/10.1080/17470218.2016.1161059

Bobes, M. A., Lopera, F., Garcia, M., Díaz-Comas, L., Galan, L., & Valdes-Sosa, M. (2003). Covert matching of unfamiliar faces in a case of prosopagnosia: an ERP study. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *39*(1), 41–56. https://doi.org/10.1016/s0010-9452(08)70073-x

Bodamer, J. (1947). Die Prosop-Agnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, *179*(1), 6–53. https://doi.org/10.1007/BF00352849

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology* , *77 (Pt 3)*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Busigny, T., Graf, M., Mayer, E., & Rossion, B. (2010). Acquired prosopagnosia as a face-specific disorder: Ruling out the general visual similarity account. In *Neuropsychologia*

(Vol. 48, Issue 7, pp. 2051–2067). https://doi.org/10.1016/j.neuropsychologia.2010.03.026

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does Prosopagnosia Take the Eyes Out of Face Representations? Evidence for a Defect in Representing Diagnostic Facial Information following Brain Damage. In *Journal of Cognitive Neuroscience* (Vol. 17, Issue 10, pp. 1652–1666). https://doi.org/10.1162/089892905774597254

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208. https://doi.org/10.1016/s0042-6989(01)00002-5

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews. Neuroscience*, *6*(8), 641–651. https://doi.org/10.1038/nrn1724

Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–1. https://doi.org/10.1167/13.10.1

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1803.11175

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014a). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014b). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570.

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and

time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187. https://doi.org/10.1002/hbm.460020306

Cohen, A. L., Soussand, L., Corrow, S. L., Martinaud, O., Barton, J. J. S., & Fox, M. D. (2019). Looking beyond the face area: lesion network mapping of prosopagnosia. *Brain: A Journal of Neurology*, *142*(12), 3975–3990. https://doi.org/10.1093/brain/awz332

Crawford, J. R., & Garthwaite, P. H. (2012). Single-case research in neuropsychology: a comparison of five forms of t-test for comparing a case to controls. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(8), 1009–1016. https://doi.org/10.1016/j.cortex.2011.06.021

Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist*, *12*(4), 482–486. https://doi.org/10.1076/clin.12.4.482.7241

Dalrymple, K. A., Oruç, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., Handy, T. C., & Barton, J. J. S. (2011). The anatomic basis of the right face-selective N170 IN acquired prosopagnosia: a combined ERP/fMRI study. *Neuropsychologia*, *49*(9), 2553–2563. https://doi.org/10.1016/j.neuropsychologia.2011.05.003

di Oleggio Castello, M. V., Haxby, J. V., & Ida Gobbini, M. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(45). https://doi.org/10.1073/pnas.2110474118

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), 1258. https://doi.org/10.1038/s41467-019-09239-1

Doerig, A., Kietzmann, T. C., Allen, E., Wu, Y., Naselaris, T., Kay, K., & Charest, I. (2022). Semantic scene descriptions as an objective of human vision. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2209.11737

Doerig, A., Sommers, R., Seeliger, K., Richards, B., Ismael, J., Lindsay, G., Kording, K., Konkle, T., Van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2022). The neuroconnectionist research programme. In *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/2209.03718

Dricot, L., Sorger, B., Schiltz, C., Goebel, R., & Rossion, B. (2008). The roles of "face" and "non-face" areas during individual face perception: Evidence by fMRI adaptation in a brain-damaged prosopagnosic patient. *NeuroImage*, *40*(1), 318–332. https://doi.org/10.1016/j.neuroimage.2007.11.012

Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: a window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173. https://doi.org/10.1016/j.conb.2006.03.003

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*, 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Dwivedi, K., Bonner, M. F., Cichy, R. M., & Roig, G. (2021). Unveiling functions of the visual cortex using task-specific deep neural networks. *PLoS Computational Biology*, *17*(8), e1009267. https://doi.org/10.1371/journal.pcbi.1009267

Eick, C. M., Kovács, G., Rostalski, S.-M., Röhrig, L., & Ambrus, G. G. (2020). The occipital face area is causally involved in identity-related visual-semantic associations. *Brain Structure & Function*, *225*(5), 1483–1493. https://doi.org/10.1007/s00429-020-02068-9

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain: A Journal of Neurology*, *135*(Pt 2), 542–554. https://doi.org/10.1093/brain/awr347

Faghel-Soubeyrand, S., Alink, A., Bamps, E., Gervais, R.-M., Gosselin, F., & Charest, I. (2019). The two-faces of recognition ability: better face recognizers extract different physical content from left and right sides of face stimuli. *Journal of Vision*, *19*(10), 136d – 136d. https://doi.org/10.1167/19.10.136d

Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022a). The neural code behind real-world recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022b). The neural code behind face recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Fiset, D., Blais, C., Royer, J., Richoz, A.-R., Dugas, G., & Caldara, R. (2017). Mapping the impairment in decoding static facial expressions of emotion in prosopagnosia. *Social Cognitive and Affective Neuroscience*, *12*(8), 1334–1341. https://doi.org/10.1093/scan/nsx068

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*(4), 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233. https://doi.org/10.1098/rsos.200233

Gao, X., Vuong, Q. C., & Rossion, B. (2019). The cortical face network of the prosopagnosic patient PS with fast periodic stimulation in fMRI. In *Cortex* (Vol. 119, pp. 528–542). https://doi.org/10.1016/j.cortex.2018.11.008

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. https://direct.mit.edu/jocn/article-abstract/12/3/495/3437

Golarai, G., Liberman, A., & Grill-Spector, K. (2015). Experience Shapes the Development of Neural Substrates of Face Processing in Human Ventral Temporal Cortex. *Cerebral Cortex* , *27*(2). https://doi.org/10.1093/cercor/bhv314

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face

recognition. *Neuropsychologia*, *49*(9), 2736–2745. https://doi.org/10.1016/j.neuropsychologia.2011.05.025

Graumann, M., Ciuffi, C., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). The spatiotemporal neural dynamics of object location representations in the human brain. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01302-0

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. In *eLife* (Vol. 7). https://doi.org/10.7554/elife.32816

Hendel, R. K., Starrfelt, R., & Gerlach, C. (2019). The good, the bad, and the average: Characterizing the relationship between face and object processing across the face recognition spectrum. *Neuropsychologia*, *124*, 274–284. https://doi.org/10.1016/j.neuropsychologia.2018.11.016

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*(5), 688–701. https://doi.org/10.1111/j.1469-8986.2004.00196.x

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial

expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *176*(2), 356–373. https://doi.org/10.1007/s00221-006-0621-5

Jiahui, G., Feilong, M., di Oleggio Castello, M. V., Nastase, S. A., Haxby, J. V., & Ida Gobbini, M. (n.d.). *Modeling naturalistic face processing in humans with deep convolutional neural networks*. https://doi.org/10.1101/2021.11.17.469009

Jiahui, G., Feilong, M., di Oleggio Castello, M. V., Nastase, S. A., Haxby, J. V., & Ida Gobbini, M. (2022). Modeling naturalistic face processing in humans with deep convolutional neural networks. In *bioRxiv* (p. 2021.11.17.469009). https://doi.org/10.1101/2021.11.17.469009

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), E6418–E6427. https://doi.org/10.1073/pnas.1802246115

Johnston, P., Overell, A., Kaufman, J., Robinson, J., & Young, A. W. (2016). Expectations about person identity modulate the face-sensitive N170. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *85*, 54–64. https://doi.org/10.1016/j.cortex.2016.10.002

Jonas, J., Rossion, B., Krieg, J., Koessler, L., Colnat-Coulbois, S., Vespignani, H., Jacques, C., Vignal, J.-P., Brissart, H., & Maillard, L. (2014). Intracerebral electrical stimulation of a face-selective area in the right inferior occipital cortex impairs individual face discrimination. *NeuroImage*, *99*, 487–497. https://doi.org/10.1016/j.neuroimage.2014.06.017

Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 861–878. https://doi.org/10.3758/s13415-013-0234-y

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

https://www.ncbi.nlm.nih.gov/pubmed/9151747

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kriegeskorte, N., & Diedrichsen, J. (2016). Inferring brain-computational mechanisms with models of activity measurements. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1705). https://doi.org/10.1098/rstb.2016.0278

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20600–20605. https://doi.org/10.1073/pnas.0705654104

Kriegeskorte, N., & Kievit, R. A. (8/2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., &

Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/s1364-6613(00)01560-6

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Liu-Shuang, J., Torfs, K., & Rossion, B. (2016a). An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia*, *83*, 100–113. https://doi.org/10.1016/j.neuropsychologia.2015.08.023

Liu-Shuang, J., Torfs, K., & Rossion, B. (2016b). An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia*, *83*, 100–113. https://doi.org/10.1016/j.neuropsychologia.2015.08.023

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level

categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38), E9015–E9024. https://doi.org/10.1073/pnas.1719616115

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. In *Electroencephalography and Clinical Neurophysiology* (Vol. 75, Issue 6, pp. 528–542). https://doi.org/10.1016/0013-4694(90)90139-b

Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: the brains behind superior memory. *Nature Neuroscience*, *6*(1), 90–95. https://doi.org/10.1038/nn988

McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *12*(1), 76–82. https://doi.org/10.1016/s0010-9452(76)80033-0

McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the*. https://www.pnas.org/content/99/3/1645.short

Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N., & Kietzmann, T. C. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(8). https://doi.org/10.1073/pnas.2011417118

Muukkonen, I., Ölander, K., Numminen, J., & Salmela, V. R. (2020). Spatio-temporal dynamics of face perception. *NeuroImage*, *209*, 116531. https://doi.org/10.1016/j.neuroimage.2020.116531

Naselaris, T., Allen, E., & Kay, K. (2021). Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences*, *40*, 45–51. https://doi.org/10.1016/j.cobeha.2020.12.008

Nemrodov, D., Niemeier, M., Mok, J. N. Y., & Nestor, A. (2016). The time course of individual face recognition: A pattern analysis of ERP signals. *NeuroImage*, *132*, 469–476.

https://doi.org/10.1016/j.neuroimage.2016.03.006

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9998–10003. https://doi.org/10.1073/pnas.1102433108

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology: CB*, *17*(18), 1568–1573. https://doi.org/10.1016/j.cub.2007.07.063

Popal, H., Wang, Y., & Olson, I. R. (2019). A Guide to Representational Similarity Analysis for Social Neuroscience. *Social Cognitive and Affective Neuroscience*, *14*(11), 1243–1253. https://doi.org/10.1093/scan/nsz099

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. In *Nature Neuroscience* (Vol. 24, Issue 11, pp. 1628–1636). https://doi.org/10.1038/s41593-021-00921-6

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421. https://doi.org/10.1016/s1364-6613(02)01976-9

Ramon, M., Busigny, T., Gosselin, F., & Rossion, B. (2016). All new kids on the block? Impaired holistic processing of personally familiar faces in a kindergarten teacher with acquired prosopagnosia. *Visual Cognition*, *24*(5-6), 321–355. https://doi.org/10.1080/13506285.2016.1273985

Richoz, A.-R., Jack, R. E., Garrod, O. G. B., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex; a Journal Devoted to the Study of the*

Nervous System and Behavior, *65*, 50–64. https://doi.org/10.1016/j.cortex.2014.11.015

Riddoch, J. M., & Humphreys, G. W. (2022). *BORB: Birmingham Object Recognition Battery*. Psychology Press. https://play.google.com/store/books/details?id=kLhmEAAAQBAJ

Rosenthal, G., Tanzer, M., Simony, E., Hasson, U., Behrmann, M., & Avidan, G. (2017). Altered topology of neural circuits in congenital prosopagnosia. *eLife*, *6*. https://doi.org/10.7554/eLife.25069

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *NeuroImage*, *40*(2), 423–426. https://doi.org/10.1016/j.neuroimage.2007.10.047

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience* , *6*(2), 258–307. https://doi.org/10.2741/E706

Rossion, B. (2018). Damasio's error - Prosopagnosia with intact within-category object recognition. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12162

Rossion, B. (2022a). Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part I: Function. In *Neuropsychologia* (Vol. 173, p. 108278). https://doi.org/10.1016/j.neuropsychologia.2022.108278

Rossion, B. (2022b). Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part II: Neural basis. *Neuropsychologia*, 108279. https://doi.org/10.1016/j.neuropsychologia.2022.108279

Rossion, B., Caldara, R., Seghier, M., Schuller, A., Lazeyras, F., & Mayer, E. (2003). A network of occipito temporal face sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain: A Journal of Neurology*, *126*(11), 2381–2395. https://doi.org/10.1093/brain/awg241

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Schiltz, C., Sorger, B., Caldara, R., Ahmed, F., Mayer, E., Goebel, R., & Rossion, B. (2006).

Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. *Cerebral Cortex* , *16*(4), 574–586. https://doi.org/10.1093/cercor/bhj005

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *80*, 141–153. https://doi.org/10.1016/j.cortex.2015.11.001

Schyns, P. G., Snoek, L., & Daube, C. (2022). Degrees of algorithmic equivalence between the brain and its DNN models. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.09.003

Simon, S. R., Khateb, A., Darque, A., Lazeyras, F., Mayer, E., & Pegna, A. J. (2011). When the brain remembers, but the patient doesn't: Converging fMRI and EEG evidence for covert recognition in a case of prosopagnosia. In *Cortex* (Vol. 47, Issue 7, pp. 825–838). https://doi.org/10.1016/j.cortex.2010.07.009

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1409.1556

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *NeuroImage*, *35*(2), 836–852. https://doi.org/10.1016/j.neuroimage.2006.09.051

Soulos, & Isik. (n.d.). Disentangled face representations in deep generative models and the human brain. *NeurIPS 2020 Workshop SVRHM*. https://openreview.net/forum?id=ME5Uh_tyld5

Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, *18*(9), 1488–1497. https://doi.org/10.1162/jocn.2006.18.9.1488

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019). Use of face information varies systematically from developmental

prosopagnosics to super-recognizers. *Psychological Science*, *30*(2), 300–308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338

Towler, J., & Eimer, M. (2012). Electrophysiological studies of face processing in developmental prosopagnosia: neuropsychological and neurodevelopmental perspectives. *Cognitive Neuropsychology*, *29*(5-6), 503–529. https://doi.org/10.1080/02643294.2012.716757

Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). FFA and OFA Encode Distinct Types of Face Identity Information. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *41*(9), 1952–1969. https://doi.org/10.1523/JNEUROSCI.1449-20.2020

Vida, M. D., Nestor, A., Plaut, D. C., & Behrmann, M. (2017). Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 388–393. https://doi.org/10.1073/pnas.1614763114

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain: A Journal of Neurology*, *107 (Pt 3)*, 829–854. https://doi.org/10.1093/brain/107.3.829

White, D., & Mike Burton, A. (2022). Individual differences and the multidimensional nature of face perception. In *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-022-00041-3

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019a). A Robust Neural Index of High Face Familiarity. *Psychological Science*, *30*(2), 261–272. https://doi.org/10.1177/0956797618813572

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young,

A. W. (2019b). A Robust Neural Index of High Face Familiarity. *Psychological Science*, *30*(2), 261–272. https://doi.org/10.1177/0956797618813572

Yan, X., & Rossion, B. (2020). A robust neural familiar face recognition response in a dynamic (periodic) stream of unfamiliar faces. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *132*, 281–295. https://doi.org/10.1016/j.cortex.2020.08.016

Zadelaar, J. N., Weeda, W. D., Waldorp, L. J., Van Duijvenvoorde, A. C. K., Blankenstein, N. E., & Huizenga, H. M. (2019). Are individual differences quantitative or qualitative? An integrated behavioral and fMRI MIMIC approach. *NeuroImage*, *202*, 116058. https://doi.org/10.1016/j.neuroimage.2019.116058

Chapitre 7 – Discussion Générale

La présente thèse s'intéresse aux mécanismes perceptifs et cérébraux expliquant les différences individuelles en reconnaissance faciale. Nous avons montré, dans les articles 1 à 3, que des variations perceptives dans le contenu visuel des représentations faciales sous-tendent des différences dans cette habileté (e.g. surutilisation des détails grossiers dans l'article 1). Ces différences expliquent autant les variations d'habileté chez des individus sains (article 3) que chez des individus provenant de populations cliniques (article 1). Dans l'article 3, nous avons même pu explorer, pour la première fois, des variations dans le contenu visuel d'un même individu à travers différentes tâches de reconnaissances faciales. Ceci nous a permis de faire la découverte étonnante que plus un individu est habile, plus le contenu de ses représentations faciales est similaire à travers les tâches. Nous avons également eu la chance de pouvoir caractériser pour la première fois les computations cérébrales d'individus avec des variations extrêmes d'habileté perceptive : des *super-recognisers* dans le top 2%, aux prosopagnosiques avec une perte de cette habileté (article 4 et 5, respectivement). L'apprentissage automatique appliqué à l'EEG nous a permis de prédire l'habileté en reconnaissance faciale des *super-recognisers,* d'individus typiques (article 4)*,* ainsi que les déficits de reconnaissance spécifiques chez un patient prosopagnosique (article 5) — et ce seulement à partir de quelques secondes de signaux cérébraux en EEG. Un apport important de cette thèse a aussi été de caractériser la relation entre les computations des modèles d'apprentissage profond, le cerveau humain, et l'habileté en reconnaissance faciale. Nous avons ainsi démontré que les computations visuelles et sémantiques (modèles de reconnaissance d'objets et langagier) varient de façon décroissante dans le cerveau des individus *super-recognisers* à un individu à l'extrême opposé du spectre d'habileté, le patient prosopagnosique PS. Ces découvertes constituent les premières démonstrations que des computations au-delà de la vision sont associées à la capacité de reconnaissance faciale. Après avoir interprété et discuté succinctement de l'impact de ces

résultats dans les prochaines sections, nous proposerons des avenues de recherches futures qui compléteront le travail que nous avons amorcé dans cette thèse.

7.1 Variations du contenu représentationnel

Notre cerveau associe nécessairement de l'information visuelle spécifique à tout type d'information reconnue visuellement. Comme nous avons vu dans l'introduction, ce contenu représentationnel peut être exprimé de façon précise sur différentes dimensions physiques d'une image comme la position, les fréquences spatiales et l'orientation de l'information, puisque ces dimensions sont essentielles aux premières transformations des objets visuels par le cerveau humain. Avec du recul, il semble ainsi assez évidemment que ces dimensions visuelles soient associées à des différences significatives dans notre habileté individuelle à reconnaître les visages (Bukach et al., 2008; Duncan et al., 2019; Dunn et al., 2022; Faghel-Soubeyrand, Dupuis-Roy, et al., 2019; Fiset et al., 2017; Pachai et al., 2013; Rossion et al., 2009; Royer et al., 2018; Tardif et al., 2019a). Les articles 1, 2 et 3 de cette thèse font suite à ces travaux. Ils ajoutent chacun une composante unique et jamais explorée préalablement dans la lignée de ces travaux que nous résumerons rapidement dans les deux prochaines sections. Nous présenterons ensuite deux avenues expérimentales futures qui permettraient de révéler ce contenu à partir de données d'imagerie cérébrale; l'une riche et complète et, l'autre très efficace.

7.1.1 Représentations visuelles anormales associées à la confusion de l'expression faciale perçue

L'article 1 a révélé une association entre contenu visuel et habileté en reconnaissance des expressions faciales émotionnelles chez une population avec un diagnostic de schizophrénie et d'anxiété sociale (SZ&SAD). Les mécanismes révélés — sous-utilisation de la région des yeux, sur-utilisation des basses par rapport à celle des hautes fréquences spatiales —, au-delà d'être une

évidence supplémentaire de ce lien entre habileté et contenu, fournissent également une compréhension unique des processus cognitifs et sociaux chez cette population. Il s'agit, tout d'abord, des premiers mécanismes perceptifs proposés pour cette population. Seulement 2 études ont été, à notre connaissance, publiées avec cette population en reconnaissance faciale (Achim et al., 2013; Lecomte et al., 2019) et ces études n'avaient recueilli aucune donnée permettant d'inférer des mécanismes perceptifs chez les individus SZ&SAD. Nos résultats font, ensuite, échos à une littérature importante sur la perception visuelle des individus sur le spectre de la schizophrénie (voir introduction de l'article pour une description exhaustive; voir aussi (Butler et al., 2008; Couture et al., 2006; Green et al., 2015; Kogata & Iidaka, 2018). Certains des corrélats perceptifs énoncés dans cet article — déficits dans le traitement précoce (magnocellulaire), déficits de sensibilité aux contrastes des basses FS (Brenner et al., 2009; Butler & Javitt, 2005; Kim et al., 2015) — sont particulièrement intéressants dans le contexte de nos résultats. En effet, ils suggèrent que les déficits de reconnaissance des visages des individus schizophrènes proviennent d'une interaction entre i) une faible sensibilité aux contrastes des basses fréquences spatiales précoce (donc faible fidélité de cette information) et ii) d'un biais de *sur-utilisation* de cette information de basses FS pour les représentations faciales. Nos résultats montrent en effet une confusion importante de cette information grossière chez les individus SZ&SAD, tout particulièrement de celle provenant de la région des yeux des visages expressifs. Nous avons proposé que ces biais perceptifs interagissent avec des biais pour des émotions négatives (i.e. particulièrement la colère et la peur; (J. Huang et al., 2011; Norton et al., 2009; Premkumar et al., 2008) pour créer des biais d'interprétation faciale (et une performance déficitaire). En effet, pris dans leur ensemble, nos résultats suggèrent que les individus SZ&SAD, même s'ils portent possiblement attention à la région des yeux, ont des représentations visuelles biaisées (en trop basses FS) qui les orientent à faussement reconnaître des expressions faciales négatives. Ce genre d'interaction entre symptômes fonctionnels et perceptifs peut évidemment avoir des effets pervers à long terme chez ces individus au niveau clinique. À ce propos, et de façon notable, cet article décrivait à notre connaissance la première analyse systématique des représentations visuelles responsables de la *confusion* des représentations faciales

(i.e. ici d'expression des émotions). Comme nous venons de le voir, ce genre d'analyses basées sur les biais de réponses (plutôt que sur l'exactitude des réponses comme dans la plupart des analyses psychophysiques; e.g., (Chauvin et al., 2005; Gosselin & Schyns, 2001), pourront s'avérer particulièrement informatives pour comprendre des populations avec des déficits perceptifs tels que ceux des populations sur le spectre la schizophrénie, mais aussi sur le spectre de l'autisme (Adolphs et al., 2008; Alink & Charest, 2020; Caplette et al., 2016; Oruc et al., 2018; Robertson & Baron-Cohen, 2017). Pour ce faire, par contre, il faudra idéalement augmenter de façon drastique la quantité d'essais enregistrés avec les méthodes de *reverse correlations* comme *Bubbles*. Même si nous avons un nombre élevé d'essais dans cet article (~30k, 1k par individus), ce genre d'analyse requiert en effet beaucoup plus d'essais que les images de classifications typiques étant donné le nombre plus faible de confusions (ou réponses incorrectes) que de réponses correctes (un ratio de 1 pour 3, en supposant que la technique d'ajustement de la performance QUEST ait fonctionné adéquatement).

Les effets trouvés montrent également que le diagnostic comorbide de SZ avec un trouble d'anxiété sociale (*Social Anxiety Disorder*, SAD) a un impact qui semble plus grand que celui de SZ ou de SAD séparément. Il faudra par contre répliquer ces résultats avec un échantillon comprenant non seulement des SZ&SAD et des neurotypiques mais également des groupes d'individus SZ et SAD.

Finalement, nos résultats ont également des applications directes au niveau clinique. Nos collaborateurs (laboratoire du Prof. Tania Lecomte) testent actuellement un entraînement perceptif en reconnaissance des expressions faciales chez des individus SZ&SAD qui sont basés sur une induction de l'information optimale de l'information faciale pour cette tâche. L'idée, simple, est similaire à celle utilisée dans (Faghel-Soubeyrand, Dupuis-Roy, et al., 2019) en révélant l'information diagnostic qu'on veut que l'individu utilise de façon implicite. Si cette technique s'avère efficace, elle pourra être implémentée simplement en demandant aux patients de compléter des entraînements à l'ordinateur dans le confort de leurs domiciles.

7.1.2 Variations du contenu représentationnel dans la population générale

Les différentes études ayant liées l'habileté en reconnaissance faciale au contenu représentationnel (Bukach et al., 2008; Duncan et al., 2019; Faghel-Soubeyrand, Dupuis-Roy, et al., 2019; Fiset et al., 2017; Pachai et al., 2013; Rossion et al., 2009; Royer et al., 2018; Tardif et al., 2019a), de même que ceux qui ont révélés le contenu représentationnel en général (voir introduction), l'ont fait en révélant soit une ou deux dimensions du contenu représentationnel (i.e. soit la position spatiale et la fréquence spatiale dans l'article 1, par exemple). Une des raisons pour ce choix est que les techniques utilisées (e.g. *Bubbles*) sont particulièrement gourmandes en temps et en ressources. De plus, aucune de ces études — ni aucune autre étude en reconnaissance faciale — n'a pu révéler les différences systématiques de stratégies d'un même individu à différentes tâches de reconnaissance faciale. Comme nous l'avons vu, ce dernier point est important puisqu'il nous empêche de répondre à des questions importantes en reconnaissance faciales: un individu habile utilise-t-il des stratégies similaires ou dissimilaires à travers ces tâches?

L'article 2 et 3 ont répondus tour à tour à ces problématiques en utilisant une nouvelle technique de *reverse correlation* i) plus efficace en ce sens qu'elle requiert moins d'essais que d'autres techniques de reverse correlation comme *Bubbles* et ii) capable de révéler simultanément l'information de position spatiale, de fréquences spatiales et d'orientation. Même si cette technique a été utilisée deux ans plus tôt par ses développeurs (Alink & Charest, 2020), l'article 2 est, en fait, à la date de publication de cette thèse, la seule étude décrivant l'utilité spécifique de ces trois attributs visuels pendant une tâche de reconnaissance visuelle (Alink et Charest ayant seulement corrélé le contenu de FS avec les traits autistiques). Le simple fait d'avoir sondé l'information d'orientation est, en soit, rare dans des tâches de reconnaissance faciale (e.g. Goffaux et al., 2016; Goffaux & Dakin, 2010; Jacques et al., 2014; Pachai et al., 2013). L'étude se rapprochant le plus de la nôtre est celle de Duncan et coll. (2017). Ces auteurs avaient associé l'information d'orientation et l'information de position avec la méthode des bulles d'orientations de façon très indirecte, en corrélant l'utilisation d'information à travers deux tâches séparées. Plus encore, ils n'avaient pas associé (directement ou

indirectement) *les trois dimensions* mentionnées plus tôt. Nos résultats sont ainsi intéressants puisqu'ils nous ont permis d'associer directement ces informations à la reconnaissance faciale de l'âge, une tâche qui avait à peine été étudiée. Nous avons utilisé cet exemple pour montrer l'utilité d'une telle méthode pour comprendre le contenu utilisé par des observateurs humains pendant une tâche visuelle : nous avons découvert que l'information provenant des yeux (position spatiale), de détails moyens (fréquences spatiales) et d'orientation horizontale (orientation) aident à la reconnaissance faciale de l'âge par des observateurs humains.

L'article trois, par contre, touchait plus directement notre question principale, i.e. *quels mécanismes expliquent que certains individus sont meilleurs que d'autres à reconnaître les visages?* Spécifiquement, nous avons répondu à la question : notre performance se généralise-t-elle d'une tâche de reconnaissance faciale à l'autre? À ce jour, une seule étude à notre connaissance (Oruc et al., 2018) s'est penchée sur cette question. Les auteurs ont observé une absence de corrélation entre l'habileté à identifier des visages et à catégoriser des émotions exprimées facialement. Dans l'article 3, nous répliquons et étendons ces résultats en mesurant l'association entre la performance dans trois tâches de reconnaissance faciale : la discrimination des émotions et l'identification de visages, tout comme Oruc et al., ainsi que la discrimination du sexe des visages. Nous montrons une association entre le discrimination du sexe et la reconnaissance des expressions, ainsi qu'entre le discrimination du sexe et l'identification des visages, mais pas, ou peu, de corrélation entre la reconnaissance des expressions et de l'identité des visages.

La question principale de cette étude était de comprendre *comment* des individus font pour être performants à différentes tâches en reconnaissance faciale. Aucune étude n'avait, jusqu'à maintenant, abordé cette question à notre connaissance. Nos résultats en *reverse correlation* ont été obtenus à partir d'un ensemble important de données recueillies sur 120 participants. Ils indiquent que plus un individu est habile en reconnaissance, plus cet individu s'appuie sur des contenus représentationnels similaires entre les tâches. Nous pensons qu'un tel mécanisme s'explique par un compromis entre notre appareil cognitif/cérébral limité, les buts diversifiés et dynamiques de notre

système visuel et les propriétés physiques des stimuli faciaux qui sont requis pour extraire ces différentes informations.

En somme, les différents résultats obtenus dans les articles 1 à 3 montrent que le construit de « contenu représentationnel » (Schyns et al. 2020) permet d'expliquer plusieurs facettes importantes de la cognition humaine : des phénomènes sociaux, des différences individuelles dans des tâches visuelles de la vie de tous les jours, et, évidemment, de leurs effets sur l'habileté perceptive à travers la population.

7.1.3 Perspectives futures sur le contenu représentationnel

7.1.3.1 Lien entre contenu représentationnel et dynamiques cérébrales

Le principe des techniques de *reverse corrélation* ne se limite pas à relier des réponses de catégorisation comportementales aux paramètres physiques manipulés expérimentalement. Cette technique peut être utilisée de façon complémentaire à l'imagerie cérébrale pour révéler l'information spécifique représentée dans le temps (c.-à-d. avec la M/EEG; (Caplette et al., 2020; Duan et al., 2018; Ince et al., 2016; Rousselet et al., 2014; Schyns et al., 2003, 2020; M. L. Smith et al., 2009, 2012; M. L. Smith, Gosselin, & Schyns, 2004; M. L. Smith, Gosselin, Cottrell, et al., 2004; Zhan et al., 2019) ou dans différentes régions cérébrales (c.-à-d. avec l'Imagerie par Résonance Magnétique fonctionnelle, IRMf; e.g. Smith et al., 2008). Dans une des premières études ayant utilisé une telle approche, Smith, Gosselin & Schyns, (2004) ont échantillonné aléatoirement l'information spatiale d'un visage à l'aide de la technique *Bubbles* (Gosselin & Schyns, 2001) tout en mesurant simultanément le signal électroencéphalographique des observateurs. En corrélant cette information présentée aux différents essais aux variations de potentiels électriques des électrodes, les auteurs ont révélé le décours temporel du traitement cortical d'une dimension (la position spatiale) du contenu des représentations perceptives de deux tâches de reconnaissance faciale. Ils ont ainsi mis en évidence deux types de processus liés aux représentations perceptives. Le premier processus précoce (~170 ms) provenait des régions occipito-temporales et était associé au traitement de l'oeil controlatéral du visage présenté du point de vue du participant, et ce peu importe la tâche

effectuée. Ce processus automatique était suivi d'un traitement spécifique au type de catégorisation effectué, vers 300 ms : les deux yeux pour la discrimination du sexe des visages et la bouche pour la discrimination de deux expressions faciales. Des études subséquentes ont peaufiné la description de ces processus temporels en montrant que le traitement de l'information visuelle lié à la tâche (l'information diagnostique) s'accentue dans le gyrus fusiforme, et ce de façon parallèle à la réduction de l'information non-diagnostique dans le cortex occipital (Duan et al., 2018; Issa & DiCarlo, 2012; Zhan et al., 2019). Le lien approfondi entre le décours temporel cérébral et le contenu représentationnel permettrait d'avoir une compréhension précise des différences d'habileté perceptive *directement liées à la tâche* (e.g. identification faciale), chose qui n'a jamais été fais à ce jours (en incluant les articles de cette thèse). Nous avons, pendant les années d'études doctorales, enregistré des données EEG haute-densité sur les mêmes 16 *super-recognisers* décrits dans l'article 4 pendant i) l'apprentissage de 8 nouvelles identités faciales et ii) pendant qu'on leurs demandait de reconnaître ces identités dans un contexte de *reverse correlation* identique à celui décrit dans l'article 3 (orientation, FS, position avec des ondelettes de Gabors). Malheureusement, nous n'avons pas encore eu le temps d'analyser ces jeux de données extrêmement riches. Quand nous le ferons, nous nous attendons à pouvoir reconstruire le décours temporel du contenu représentationnel dans le cerveau de *super-recognisers* pendant qu'ils traitent l'information d'identité faciale. Ceci sera la première description du lien entre stratégie visuelles de ces individus (Dunn et al., 2022; Tardif et al., 2019a) et leurs dynamiques cérébrales. Des résultats préliminaires ayant relié cette technique psychophysique au décours temporel de l'information multivarié (MVPA) dans le cerveau a déjà été présentés dans un article de conférence à la conférence Cognitive Computational Neuroscience (CCN) et est inclut en annexe I de cette thèse (Faghel-Soubeyrand, Alink, Bamps, Gosselin, et al., 2019).

7.1.3.2 Révéler le contenu représentationnel cérébral efficacement avec le Frequency Tagging

Les techniques de *reverse-correlation* et de MVPA sont coûteuses en temps et en argent pour l'expérimentateur et donc peu appropriées, à long terme, pour l'étude des différences individuelles

où les échantillons doivent être de taille importante. De nouvelles méthodes permettant de sonder les représentations visuelles individuelles plus rapidement sont donc nécessaires. Nous avons commencé à développer et à tester une telle méthode. Celle-ci est basée sur le *Frequency Tagging* (voir Norcia et al., 2015, pour un survol de la litérature), une méthode où la stimulation visuelle rapide et répétée à une fréquence précise induit des oscillations cérébrales qui peuvent être facilement captées par une analyse Fourier des données obtenues en imagerie cérébrale M/EEG. Cette technique exploite un résultat décrit par Regan en 1966 (voir aussi Adrian & Matthews, 1934) : si une lumière clignote à une fréquence relativement rapide et constante *f*1, les populations de neurones de l'observateur répondant au stimulus visuel tendront à se synchroniser à cette fréquence de stimulation et les multiples entiers de ses harmoniques (c.-à-d. 2**f*1, 3**f*1, 4**f1*, etc.). Cet entraînement neuronal est appelé *Steady-State Visual Evoked Potentials* (SSVEP) à cause de la stabilité de la phase et de l'amplitude du signal cérébral enregistré aux fréquences induites par la stimulation. L'utilisation la plus courante de la technique de *frequency-tagging* permet de suivre le traitement d'un stimulus ou d'une partie d'un stimulus chez un individu en fonction de divers paramètres expérimentaux. Cette technique a ainsi permis de répondre à des questions variées portant sur la conscience visuelle (Parkkonen et al., 2008), l'attention spatiale (e.g. Morgan et al., 1996), l'attention basée sur les objets (*object-based attention*, e.g. Baldauff & Desimone, 2014) et l'attention basée sur les attributs des objets (*feature-based attention,* Andersen et al., 2009; Andersen & Hillyard, 2014;Andersen et al., 2008). En demandant à des participants de porter attention à l'un des deux objets présentés simultanément dans les deux hémichamps visuels et en les faisant osciller, respectivement, à deux fréquences de 8.5 et 12 Hz, Morgan et coll. (1996) ont, par exemple, été capables de quantifier l'allocation de l'attention spatiale à chacun des stimuli. Notons qu'il est difficile, voire impossible, de mesurer la contribution *respective* de différents objets *présentés simultanément* avec les méthodes plus classiques de présentation en ERP. Un expérimentateur utilisant le frequency tagging peut ainsi tirer profit du signal SSVEP pour quantifier objectivement (les SSVEPs donnent une réponse claire qui est nécessairement établie expérimentalement *a priori*) et de façon exacte (les SSVEPs ont un rapport signal sur bruit élevé) le

traitement de stimuli spécifiques présentés simultanément. Nous avons développé une technique basée sur cette idée pour mesurer le contenu représentationnel spatial chez des participants neurotypiques. Celle-ci utilise des *frequency-tags* pour chacun des trois attributs faciaux les plus importants pour la reconnaissance faciale dans diverses tâches, c.-à-d. l'oeil gauche, l'oeil droit, et la bouche (e.g. Gosselin et Schyns, 2001 ; Schyns, Bonnar & Gosselin, 2002). Pour valider cette mesure de contenu représentationnel, nous l'avons comparée à la technique de *reverse corrélation Bubbles* (Gosselin & Schyns, 2001). Une version préliminaire de ce projet a été présentée à la conférence Vision Science Society en 2017 (Faghel-Soubeyrand & Gosselin, 2017) et 2021 (Gervais et al., 2021). Nos résultats préliminaires sont encourageants pour deux raisons. La première est que nous arrivons à mesurer l'utilisation de ces attributs faciaux en un peu plus d'une minute avec un rapport signal sur bruit élevé (~1600%, voir figure 7.1b). La seconde est que ce signal *taggé* semble i) correspondre aux modulations de l'attention spatiale d'un individu et ii) permettre de mesurer l'*intégration* des attributs faciaux étiquetés au sein du cortex (Faghel-Soubeyrand & Gosselin, 2017). Ceci est, en effet, possible grâce à la capacité de cette technique de capter des *fréquences intermodulaires* (intermodular frequencies, IM; voir Norcia et al., 2015), i.e. les composantes fréquentiels non-linéaires du signal cérébral taggé qui résultent de la somme et de différences des tags, par exemple $f1 \pm f2$. Cette technique nous a ainsi permis de quantifier le rôle respectif du traitement de chaque attribut facial (par les tags fréquentiels uniques, e.g. oeil gauche par $f1$), ainsi que leur intégration dans le cerveau pendant qu'un participant résout une tâche particulière (par les fréquences IMs, e.g. intégration de l'oeil gauche et oeil droit par $f1+f2$). Nous avons pu montrer que les IM quantifient avec succès la composante attentionnelle de l'intégration cérébrale liée à la tâche : les fréquences IM associée aux deux yeux — la réponse neuronale attribuable à l'intégration des deux attributs — était plus forte lorsque les participants devaient traiter les informations des deux yeux par rapport à la condition d'attention focalisée sur un seul oeil, par exemple (Faghel-Soubeyrand & Gosselin, 2017). Cette technique est une alternative prometteuse aux techniques de *reverse correlation* qui sont nettement plus gourmandes en temps.

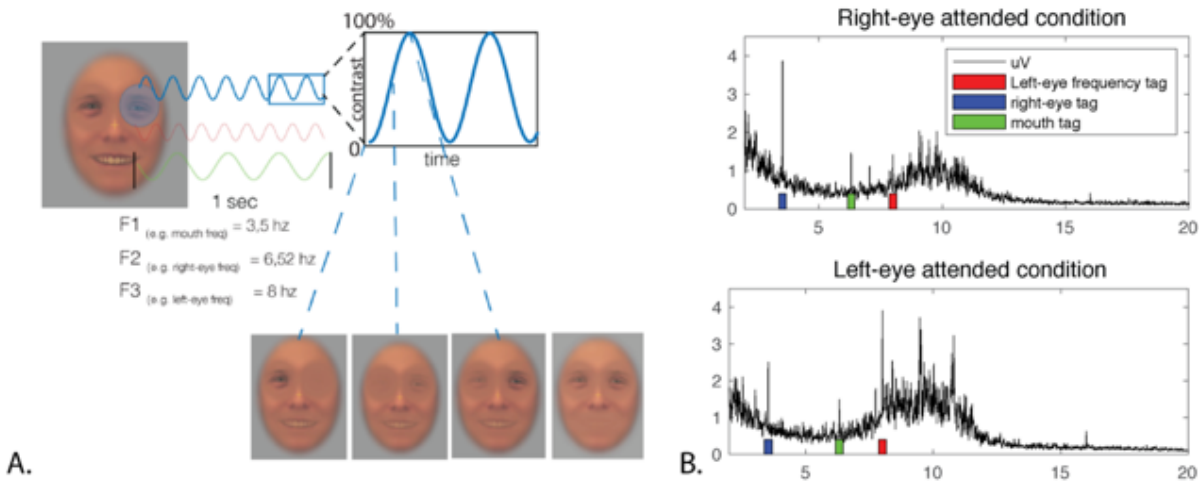


Figure 7.1 Méthode expérimentale de frequency tagging pour capturer le contenu représentationnel cérébrale. A) Le contraste de trois attributs (l'oeil gauche, l'oeil droit, la bouche) d'un visage oscillent selon des ondes sinusoidales de fréquences spécifiques (par exemple, 3,5290 Hz pour l'œil gauche, 8,000 Hz pour l'œil droit et 6,3150 Hz pour la bouche). Lors d'un essai typique, une croix de fixation centrée sur l'écran était superposée au visage de fond tandis que ces trois attributs oscillent à leurs fréquences respectives. Les caractéristiques du même visage oscillent pendant toute la durée d'un long essai de 70 secondes. B) Le spectre fréquentiel du signal cérébral tel que mesuré par l'EEG, moyenné sur les électrodes occipito-temporales après une période de 70 secondes/condition pour un sujet type. La couleur indique le trait facial marqué: rouge pour l'oeil gauche, bleu pour l'oeil droit, vert pour la bouche. Le ratio signal sur bruit de ce signal est extrêmement élevé, atteignant >16 (1600%) chez certains participants ($p<.000001$; voir (Faghel-Soubeyrand & Gosselin, 2017; Gervais et al., 2021). Ces résultats sont étonnants si on compte qu'ils ont été obtenus en un peu plus d'une minutes.

7.2 Variations des computations cérébrales

Quels sont les mécanismes cérébraux qui expliquent les différences individuelles en reconnaissance faciale? La littérature en neuroimagerie portant sur cette question (Elbich & Scherf, 2017; Herzmann et al., 2010; L. Huang et al., 2014; Kaltwasser et al., 2014; Lohse et al., 2016; Nowparast Rostami et al., 2017; Rossion et al., 2020) a utilisé des mesures indirectes de traitement tel que la sélectivité de l'activation cérébrale aux visages par rapport aux autres objets visuels (e.g. dans le FFA gauche dans Elbich & Scherf, 2017; et dans les deux FFAs dans Huang et al., 2014), et la latence des composantes cérébrales reliées au traitement facial tel que la N170 (Herzmann et al., 2010; Kaltwasser et al., 2014; Nowparast Rostami et al., 2017). Ces méthodes ne permettent pas de révéler les *mécanismes* cérébraux en-soi mais plutôt des corrélats cérébraux de la reconnaissance faciale et

de leur lien avec l'habileté. Ces corrélats ne nous renseigne pas sur la *nature* des représentations perceptives qui sous-tendent ces différences individuelles. Les articles 4 et 5 sont, à notre connaissance, les premiers à tenter d'identifier les computations cérébrales expliquant les différences individuelles en reconnaissance de visages.

L'article 4 a révélé que le cerveau des individus avec une habileté exceptionnelle en reconnaissance faciale — les *super-recognisers* — effectue des computations plus similaires à celles effectuées par des modèles d'apprentissage profonds de reconnaissance d'objets (CNNs; Krizhevsky et al., 2012; Yamins & DiCarlo, 2016) que les individus avec une habileté typique. Comment interpréter ce résultat? Nous pensons qu'il est important de considérer d'abord ce que sont ces CNNs visuels : des modèles de reconnaissance *optimaux*, capables d'effectuer des tâches de reconnaissance avec une performance souvent supérieure à celles des humains (e.g. Phillips et al., 2018). Comparés aux modèles d'observateurs idéaux développés dans l'article 3, par exemple, ces modèles i) effectuent des tâches beaucoup plus complexes, ii) sont plus robustes aux distortions d'images de bas niveau et iii) effectuent ces opérations selon un nombre de paramètre extrêmement plus grands que ces observateurs linéaires. Ces opérations, telles que discutées dans l'introduction, sont *similaires* à celles effectuées par le système *visuel* humain (voir figure 1.6b). Nous proposons que cette similarité plus grande entre le traitement cérébral des *super-recognisers* et celui des CNNs visuels proviennent des computations se rapprochant davantage de l'optimalité dans les deux cas. Spécifiquement, nos résultats montrent que des computations de niveaux moyens d'abstractions (Groen et al., 2017; Li & Bonner, 2021; Long et al., 2018), correspondant à des traitements de parties d'objets ou de visages (Güçlü & van Gerven, 2015), est le niveau où les computations cérébrales des super-recognisers sont plus optimales. Ce résultat nous semble intéressant pour plusieurs raisons. D'abord, parce que ceci constitue le premier mécanisme cérébral proposé qui pourrait expliquer en partie l'habileté exceptionnelle en reconnaissance faciale des *super-recognisers*. Il est également possible que les *super-recognisers* aient des représentations visuelles intermédiaires plus robustes qui leur permettraient d'être performants dans plusieurs tâches (voir article 3). Nous avons également révélé

le décours temporel de ce mécanisme dans le cerveau : l'habileté perceptive semble augmenter la qualité des représentations visuelles de niveaux intermédiaires (couche 4 sur 8 dans AlexNet, par exemple) entre 135 ms et 165 ms après la présentation d'un stimulus, c'est-à-dire exactement dans la fenêtre temporelle de la N170 (Bentin et al., 1996; Heisz et al., 2006; Rossion & Jacques, 2012; Shen et al., 2017). Encore une fois, ce résultat fait écho à ceux d'études ayant utilisé l'EEG avec des méthodes univariées chez des individus typiques (Herzmann et al., 2010; Kaltwasser et al., 2014; Nowparast Rostami et al., 2017) et ayant observé des différences de latences autour de cette fenêtre de traitement mais, contrairement à notre étude, sans jamais avoir caractériser la nature du processus dans le cerveau. Cet effet des représentations visuelles intermédiaires sur l'habileté se produit précisémment au moment où le «decoding» de l'habileté individuelle atteint son sommet. Notons que ce «decoding» de l'habileté à partir de l'EEG de nos participants est une première, que cet effet semble robuste (~80% précision, «cross-validée» sur des dizaines de milliers d'essais) et que le décours temporel de ce decoding suggère que des processus cérébraux divers sont impliqués, incluant évidemment des processus perceptifs visuels. De plus, ces processus visuels semblent non spécifiques aux visages. La spécificité des processus en reconnaissance faciale est encore fortement débattue (Gauthier, Tarr, et al., 2000; Grill-Spector et al., 2004; Kanwisher, 2000; Moscovitch et al., 1997). Notre étude fournit des évidences convaincantes que la capacité de reconnaissance des visages covarie avec des mécanismes cérébraux communs entre la reconnaissance des objets et la reconnaissance des visages et appuie donc les théories de la reconnaissance faciale dites «domaine-générale».

Notre résultat le plus surprenant est sans doute celui qui montre que les différences individuelles en reconnaissance de visages reposent, en partie, sur des computations cérébrales qui vont au-delà du traitement visuel, computations que l'on qualifie de sémantiques. Nos analyses ont, en effet, révélé une longue fenêtre temporelle de traitement cérébral pendant laquelle les computations effectuées par le cerveaux des *super-recognisers* semblent plus similaires aux opérations d'un modèle d'apprentissage profond langagier, entraîné à prédire des différences sémantiques provenant des

phrases et de courtes descriptions écrites (Cer et al., 2018). Comme mentionné dans l'article, ceci constitue la toute première démonstration que des computations sémantiques participent à notre capacité à reconnaître les visages. Plusieurs modèles influents en reconnaissance faciale ont une composante sémantique (Bruce & Young, 1986; Duchaine & Yovel, 2015; Haxby et al., 2000; White & Mike Burton, 2022), mais aucune étude n'avait encore corroboré empiriquement cette proposition. Nos résultats indiquent que plus un participant est habile à reconnaître les visages, plus riches ont tendance à être les liens sémantiques entre les différentes représentations visuelles dans son cerveau. Ainsi, cette capacité perceptive n'impliquerait pas seulement le traitement visuel traditionnellement étudié en reconnaissance de visages tel que présenté brièvement dans l'introduction de cette thèse — le contenu représentationnel (Gosselin & Schyns, 2001), la sélectivité cérébrale aux visages (Kanwisher, 2000) ou les dimensions du «face-space» (Dennett et al., 2012) — mais elle ferait aussi intervenir différentes dimensions sémantiques associées aux objets — e.g. leurs fonctions (Collette et al., 2016; Weisberg et al., 2007) et leurs niveaux de catégorisation (Graf & Bülthoff, 2019; J. W. Tanaka & Taylor, 1991; Jozwik et al., 2022). Nous acquérons actuellement des données comportementales à l'aide d'une expérience réalisée en ligne qui devraient nous permettre d'explorer davantage ces questions. Nous utilisons des données multidimensionnelles qui permetteront de produire des RDMs créés sur la base de similarités visuelles (e.g. couleurs, formes) et sémantiques (e.g. fonction, niveau de catégorisation). Ces RDMs comportementales seront comparés aux RDMs cérébrales des articles 4 et 5 afin de caractériser plus finement les processus sous-jacents aux effets découverts chez les super-recognisers et la patiente prosopagnosique PS.

Comment se comparent, justement, ces résultats avec ceux obtenus avec la patiente prosopagnosique PS? PS a des lésions à l'OFA droit et au FFA gauche, et démontre une prosopagnosie sévère et très spécifique à la reconnaissance faciale d'*identité*, comme l'ont montré une série de plusieurs études maintenant classiques (Rossion, 2022a). Nous avons eu la chance de tester cette femme en Octobre 2019; PS démontre, soit dit en passant, une mémoire générale exceptionnelle, une attention soutenue

exemplaire, et une joie de vivre contagieuse. Nous avons pu faire plusieurs démonstrations importantes suite à ces enregistrements. Deux peuvent être considérées comme méthodologiques. La première, rassurante, a été de montrer que les *dynamiques* cérébrales de PS en EEG haute-densité sont fiables à travers différents jours d'enregistrement quand on utilise la RSA pour s'abstraire des topographies brutes. Ceci peut paraître trivial, mais il est particulièrement difficile d'avoir un signal EEG interprétable avec un individu ayant des lésions cérébrales significatives (pour une discussion, voir : Alonso Prieto et al., 2011; Liu-Shuang et al., 2016). Ceci montre que l'utilisation de la RSA est prometteuse pour la caractérisation des dynamiques cérébrales d'autres études de cas, de cas cliniques, et des neurosciences sociales plus généralement (Popal et al., 2019). La deuxième démonstration est plus surprenante et convaincante: nous avons pu révéler les dynamiques cérébrales des déficits cérébraux spécifiques au visages seulement à partir d'enregistrement EEG. Plus précisément, nous avons pu capturer i) l'incapacité de son cerveau de discriminer l'identité des visages, et, plus surprenant encore ii) nous avons pu capturer la capacité totalement préservée de son cerveau de discriminer l'identité des non-visages (Rossion et al., 2009). Ces résultats sont étonnants puisqu'ils signifient que nous pouvons prédire, à partir d'un cerveau significativement endommagé, l'identité de différents objets, scènes et animaux *à travers le temps*. En effet, aucune étude à ce jour n'a révélé de façon convaincante les dynamiques cérébrales chez PS (Alonso Prieto et al., 2011). Ceci montre également que les résultats computationnels que nous avons obtenus avec cette patiente ne peuvent pas être dû exclusivement à des problèmes de ratio signal sur bruit, puisque des objets peuvent être décoder typiquement à partir de son signal cérébral.

Les résultats computationnels principaux ont mené à la découverte que PS démontre des computations cérébrales visuelles et des computations cérébrales sémantiques déficitaires. Ces résultats appuient et étendent d'abord ceux de l'article 4. Ils montrent que les computations cérébrales visuelles et sémantiques suivent le spectre de l'habileté en reconnaissance faciale : de la prosopagnosie, aux individus avec une habileté typique, jusqu'aux *super-recognisers* (Tardif et al., 2019a). Ils sont, tout comme ceux de l'article 4, les premiers ayant révélé la *nature* des représentations cérébrales en prosopagnosie. Si ce déficit important a été étudié depuis près de 150

ans maintenant (Corrow et al., 2016), peu d'études ont pu révéler des dynamiques cérébrales convaincantes (Alonso Prieto et al., 2011; Dalrymple et al., 2011) — et ce même en prosopagnosie développementale, e.g. voir (Corrow et al., 2016) — et aucune n'a révélé la nature des déficits cérébraux dans ce trouble perceptif. De façon intéressante, les computations touchées étaient de plus haut niveaux que celles chez les *super-recognisers* (atteignant un maximum aux couches 6 sur 8 du CNN visuel; Krizhevsky et al., 2012), et les computations cérébrales sémantiques semblaient être significativement plus touchées (tant au niveau temporel que de la magnitude des différences en comparaison avec les *super-recognisers*). Ces résultats s'alignent avec le fait que PS présente des troubles de reconnaissance associatifs, et pas ou peu de troubles perceptifs de bas niveau (Rossion, 2022a, 2022b), et que l'information d'identité est associée à une gamme d'information riche dans le système visuel (Ambrus et al., 2017; A. J. Anderson et al., 2020; Bruce & Young, 1986; Jonas et al., 2014), possiblement dans les régions du OFA/FFA.

À ce propos, nous terminions la discussion de l'article 5 sur une note exploratoire quant aux régions cérébrales qui pourraient possiblement sous-tendre ces déficits computationnels : PS a en effet des lésions dans le OFA droit, et dans le FFA gauche, et d'une partie plus antérieure du gyrus temporal médian droit (Sorger et al., 2007). Quelles régions cérébrales peuvent sous-tendre les computations sémantiques révélées dans la présente thèse? Très peu d'études en neurosciences cognitives computationnelles ont abordé cette question en date de dépôt des articles 4 et 5 (Doerig, Kietzmann, et al., 2022; Dwivedi et al., 2021; Popham et al., 2021; Visconti di Oleggio Castello et al., 2021). Deux d'entre elles (Doerig, Kietzmann, et al., 2022; Popham et al., 2021) sont particulièrement d'intérêt puisqu'elles ont utilisé des méthodes *similaires* aux nôtres. Ces études ont révélé un réseau très diffus de régions, incluant des régions du cortex inféro temporal (e.g. FFA, LOC, PPA), mais également des régions du lobe temporal antérieur (ATL), qui sont prédites par un modèle sémantique. Ces études toutefois, utilisaient des procédures expérimentales différentes de celles utilisées ici. Afin de déterminer la localisation des computations sémantiques trouvés dans les articles 4 et 5, nous avons utilisé des données fMRI provenant de Charest et al. (2014) qui ont un

espace de stimuli très similaire au nôtre (36 stimuli, avec des visages, des scènes, et des objets, en proportion similaire; voir Charest et al., 2014 pour plus de détails). Nous avons utilisé ces données fMRI (3T, avec couverture entière du cerveau) pour dériver des RDMs de 36 x 36 pour chaque voxel du cerveau, et avons ensuite associé (Ince et al., 2017) ces représentations cérébrales à celles du même modèle sémantique utilisé dans les articles 4 et 5 (GUSE; Cer et al., 2018). L'image résultant de cette analyse, montrée à la figure 7.2, est une carte cérébrale montrant les régions qui encodent les stimuli visuels de façon semblable aux computations sémantiques du modèle GUSE (régions plus chaude : plus similaires aux computations sémantiques). Notons que ces RDMs sont produites à partir des *phrases descriptives* de ces stimuli visuels. Ceci a révélé un large réseau de régions couvrant, de façon intéressante : l'OFA, le FFA, et le PPA, particulièrement dans l'hémisphère droit. Aussi observées sont des régions de plus haut niveaux telles que le posterior cingulate cortex (pCC) et des régions du lobe temporal médian (MTL). Ce réseau de régions, particulièrement le OFA et FFA, est bien sûr traditionnellement impliqué dans la perception faciale, et ces régions sont lésées chez PS. Cette analyse est intéressante puisqu'elle ajoute à l'idée que ces régions renferment un code représentationnel qui est beaucoup plus riche et diversifié (Jonas et al., 2014, 2015; Rossion, 2022b) que ce que les études classiques en reconnaissances faciales suggèrent avec les analyses univariées décrites dans l'introduction (*idem* pour les modèles standards en reconnaissance faciale, e.g. Duchaine & Yovel, 2015). Nous avons également produit une carte cérébrale de l'information visuelle de niveau intermédiaire, i.e. des computations provenant d'un modèle d'apprentissage profond de reconnaissance d'objet (AlexNet; (Krizhevsky et al. 2012). Cette analyse nous montre, de façon complémentaire, des régions visuelles de moyen niveau (i.e. V3, hV4, LOC) et de bas niveau (V1,V2). Ces analyses nous permettent ainsi d'interpréter avec un peu plus d'assises les computations visuelles, et visuo-sémantiques révélées dans l'article 4 et 5. En effet, les régions associées à ces computations visuelles et visuo-sémantiques et montrée à la figure 7.3 et 7.2 suggèrent que les computations trouvées dans les articles 4 et 5, qui sont graduellement plus élevés de PS aux *super-recognisers*, proviennent probablement d'un réseau cérébral beaucoup plus vaste que nous le pensions (Harel et al., 2010). Ce réseau *inclut* les régions classiques du réseau classique

de la perception faciale (Duchaine and Yovel 2015), mais n'y est pas circonscrit (Haxby et al., 2000, 2001). Cette conclusion s'aligne également avec le décours temporel du decoding de l'habileté dans l'article 4 : nous avions observé des différences fonctionnelles de 65 ms à près de 1 s après la présentation du stimulus. Il est certain qu'à ces latences, un réseau très étendu et diversifié de régions est à l'œuvre dans le cerveau.

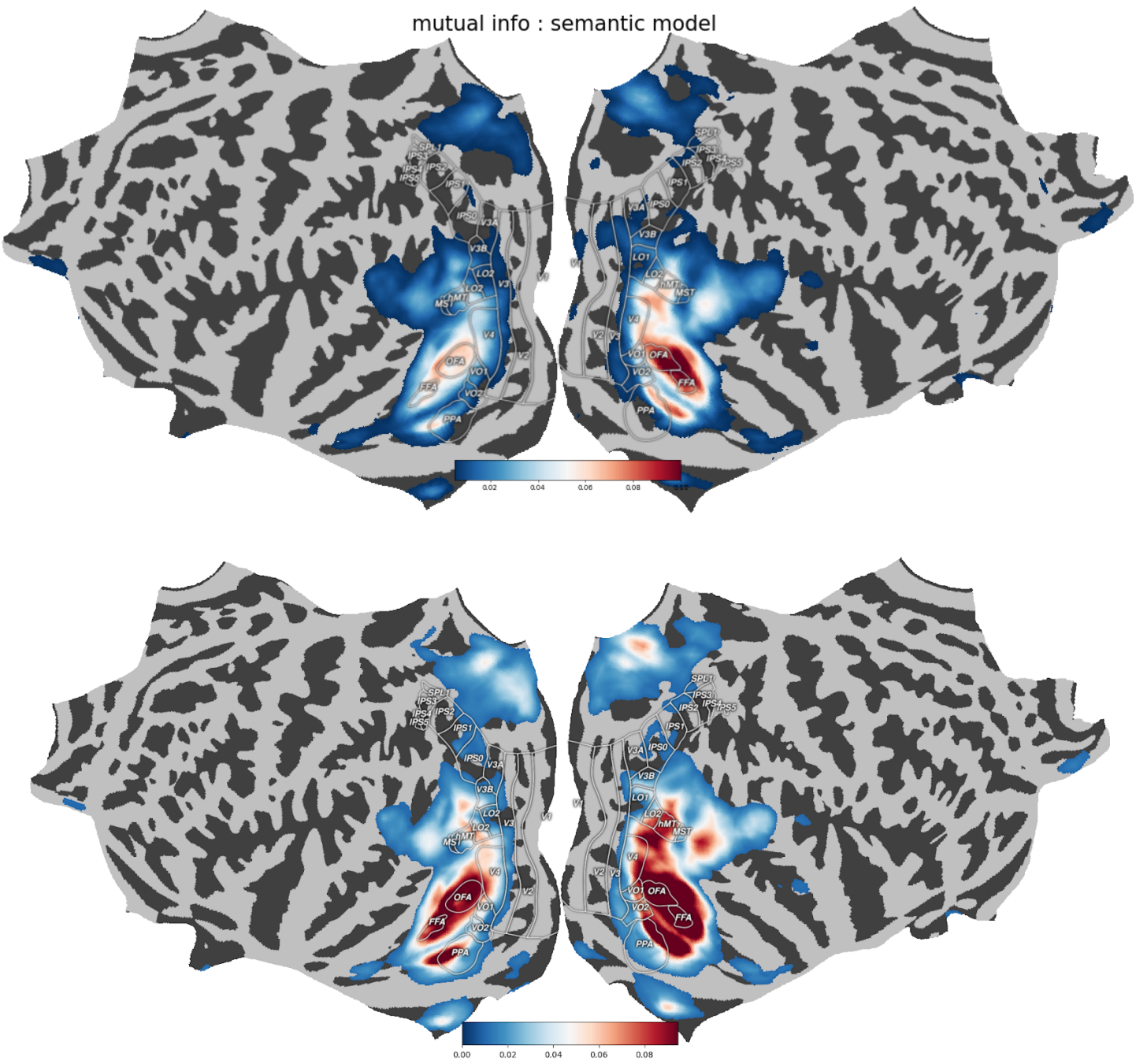


Figure 7.2. Régions cérébrales associées au traitement visuo-sémantique d'un modèle d'apprentissage profond langagier. Le modèle Google Universal Sentence Encoder (Cer et al., 2018) utilisé dans l'article 4 et 5 de cette thèse a été associé au représentations cérébrales fMRI. Spécifiquement, des données fMRI

provenant de 20 participants de (Charest et al., 2014) ont été utilisées afin de créer des RDMs par voxel (voir Charest et al., 2014). Une RDM sémantique du Google Universal Sentence Encoder (GUSE) a été produite en générant des phrases descriptives sur les stimuli utilisée dans Charest et al., (2014), et en utilisant la même méthode que celle utilisée dans les articles de cette thèse, i.e. en donnant ces descriptions en input au modèle GUSE, et en calculant la dissimilarité entre l'espace sémantique de ce modèle en réponse à ces descriptions écrites. Le traitement des différentes régions corticales — les RDMs des différents voxels à travers le cerveau — ont été associées à la RDM sémantique avec la même procédure de Mutual Information (MI) (Ince et al., 2017) que celle décrite dans l'article 4, avec des couleurs chaudes pour une plus haute similarité (MI cut-off à > 0). Cette figure montre ainsi les régions cérébrales qui ont des représentations les plus semblable à celles du modèle sémantique, notamment : le OFA, FFA, et PPA, et le LOC, mais également des régions encore plus antérieures du lobe temporal (c.f. Doerig, Kietzmann, et al., 2022). Le premier panel montre la MI entre les voxels et ce modèle sémantique, controllées pour un modèle artificiel profond de la vision (voir prochaine figure; comme décrit dans les articles 4 et 5). Le deuxième panel montre l'association simple entre les RDMs des voxels à travers le cerveau et ce modèle sémantique.

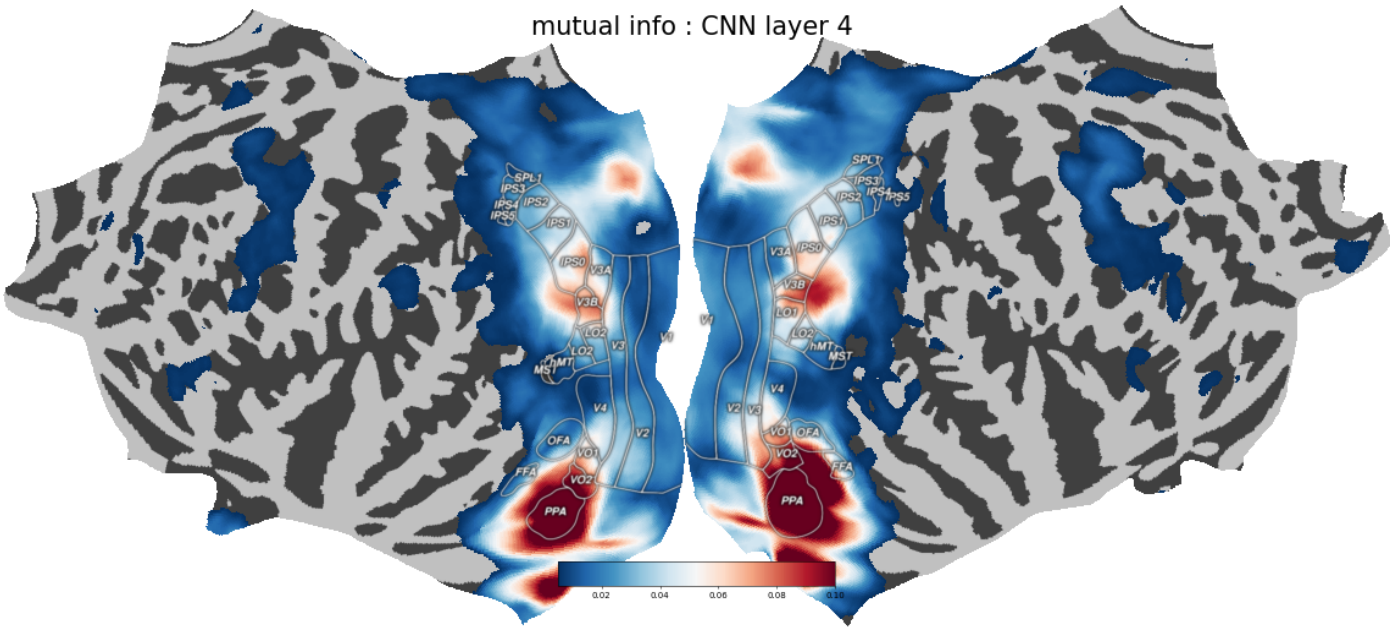


Figure 7.3. Régions cérébrales associées au traitement visuel du modèle d'un modèle visuel d'apprentissage profond de reconnaissance d'objets. Le modèle d'apprentissage profond convolutionnel de reconnaissance d'objet AlexNet (Krizhevsky et al. 2012) utilisé dans l'article 4 et 5 de cette thèse a été associé au représentations cérébrales fMRI de façcon identique à la procédure décrit à la figure 7.2, mais cette fois en utilisant la réponse multivariée des différentes couches du modèle visuel à aux *images* présentés aux participants de Charest et al., (2014). Une RDM a été crée pour chacune des 8 couches de ce modèle (voir introduction). Les résultats pour la couche intermédiare 4 du CNN sont présentés pour son importance dans

l'article 5 – un traitement similaire à ces computations visuelles intermédiaires discrimine le traitement cérébrale des *super-recognisers* vs. des *typical recognisers*. Cette figure montre des régions cérébrales qui ont des représentations les plus semblable à celles de ces computations visuelles intermédiaires, notamment : le V2-V3-hV4, LO1-LO2, VO1-VO2, et le PPA, mais également des régions encore plus postérieures du système visuel (c.f. Dwivedi et al. 2021). Cette figure montre la MI entre les voxels et le modèle AlexNet, controllées pour l'information sémantique (voir figure précédente; comme décrit dans les articles 4 et 5).

7.3 Perspectives futures générales

L'ensemble des résultats présentés dans cette thèse montre que l'habileté en reconnaissance faciale est modulée par des mécanismes riches et diversifiés, s'échelonnant à plusieurs étapes temporelles du traitement cérébral au sein d'un réseau tout aussi diversifié de régions cérébrales. Au delà des différents mécanismes perceptifs, cérébraux, et computationnels derrière l'habileté perceptive qui ont été décrits dans les articles 1 à 5, nous croyons que ces résultats ont une valeur heuristique importante pour plusieurs disciplines; en vision, en neuropsychologie, en neurosciences-cognitives et computationelles, et en intelligence artificielle.

Pour l'étude des différences individuelles en perception, ces résultats ouvrent la voie à diverses questions, expérimentations, et analyses approfondies, notamment i) sur le lien entre le contenu représentationnel et les computations cérébrales visuelles (voir aussi Daube et al., 2021; Schyns et al., 2022), ii) sur l'effet spécifique des différentes dimensions sémantiques sur notre capacité de reconnaissance, iii) sur les régions cérébrales spécifiques qui médient la relation entre computations et habileté perceptive et iiii) sur la causalité des computations cérébrales visuelles et sémantiques sur cette habileté. À ce niveau, nous avions mentionné dans l'article 5 qu'il serait intéressant d'associer la résolution spatiale de la fMRI, la résolution temporelle de l'EEG, et les computations des CNNs pour avoir le décours spatio-temporel des computations cérébrales. La carte spatio-temporelle illustrée à la figure 7.3 est une première étape pour relier les données fMRI et EEG dans ce sens. Nous avons utilisé la RSA pour associer des RDMs produites à partir de données fMRI de Charest et coll. (2014) et des RDMs produites à partir d'EEG haute-densité (non-publiées) avec la Mutual Information (Ince et al., 2017), de façon similaire à la procédure de fusion décrite dans (Cichy & Oliva, 2020). Cette méthode de fusion RSA nous a permis de produire un court vidéo qui décrit le

code représentationnel dans le temps (de -50 ms à 250 ms) et dans l'espace des voxels du cerveau humain. Quand cette carte spatio-temporelle sera associée i) aux computations des CNNs visuels et sémantiques et ii) à des mesures perceptives individuelles, nous pourrons avoir un portrait d'une précision inégalée des mécanismes cérébraux derrière l'habileté en reconnaissance faciale.

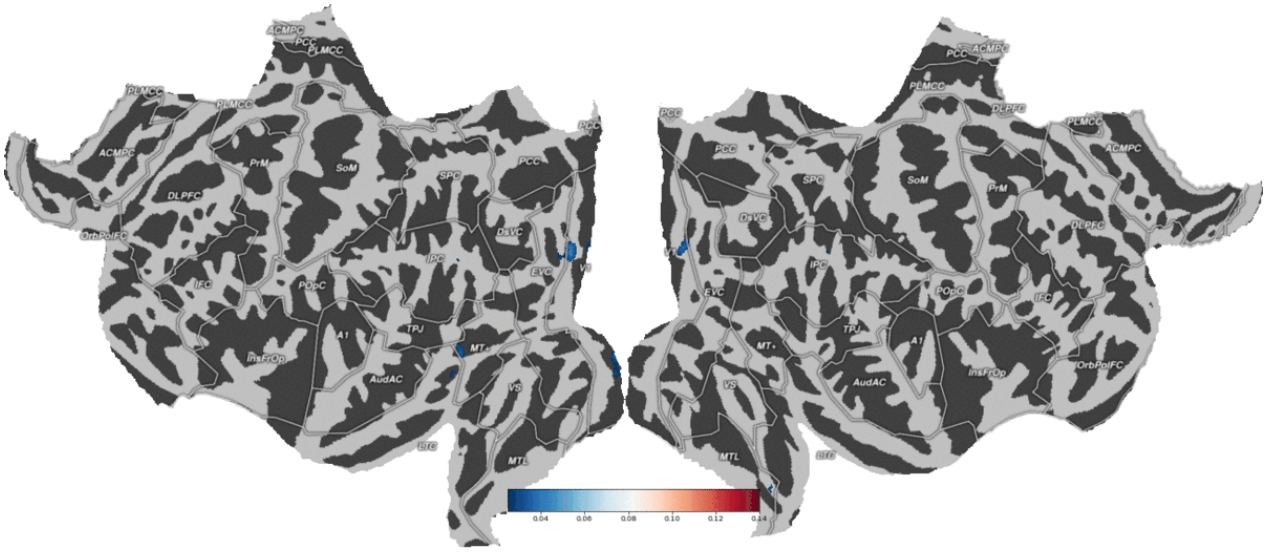


Figure 7.4. **Décours spatio-temporel du traitement visuel dans le cortex humain — les premières 250 ms de traitement**. Cette figure est un vidéo disponible au lien suivant: lien youtube ici. Nous avons associé le signal fMRI de Charest et coll. (2014) et le signal électroencéphalographique en EEG haute-densité (non-publiées) en réponse aux mêmes images en utilisant la méthode de fusion en RSA (Cichy & Oliva, 2020). La fusion de ces patrons spatio-temporels a été produite en associant la RDM de chaque voxel aux RDMs EEG à chaque moment de -50 ms à 250 ms à partir de la présentation du stimulus. La taille des RDMs en fMRI et en EEG est identique à celles décrites à la figure 7.2. Cette analyse montre une progression postérieur à antérieure de l'information visuelle dans le cerveau humain, avec des représentations entre EEG et fMRI qui corrèlent le plus vers 135 ms après la présentation du stimulus, dans le cortex inféro temporal du *ventral stream* (VS) et dans le *Medial temporal Lobe* (MTL).

Les résultats et méthodes utilisées sont aussi intéressantes pour l'étude de la neuropsychologie et des applications cliniques. Tout d'abord, les modèles de decoding développés dans l'article 4 sont directement applicables pour la détection d'individus *super-recognisers* dans la population. Nous

avons été capables de prédire à près de 80% de sensibilité cette catégorie d'individus à partir d'une seule seconde d'activité EEG. Évidemment, ces modèles devront être testés pour leur généralisation, mais la quantité de données utilisées suggère que ces modèles seraient déjà suffisamment performants pour recruter des individus avec des emplois requérant un traitement facial efficace tel que les policiers, agents de sécurité et agents de la douane.

Deux lignes générales d'entraînement perceptif sont également possibles. D'abord, nos résultats du contenu représentationnel pourront être utilisés pour développer des entraînements qui forcent l'utilisation du contenu le plus optimal chez des individus avec des troubles perceptifs, tels que chez des individus prosopagnosiques, ou sur le spectre de l'autisme et de la schizophrénie (Couture et al., 2006; Green et al., 2015; Robertson & Baron-Cohen, 2017). De telles méthodes ont été testées et semblent prometteuses (Adolphs et al., 2005; Faghel-Soubeyrand, Dupuis-Roy, et al., 2019; A. Towler et al., 2021). Une deuxième possibilité serait de créer, à partir des données cérébrales collectées à travers nos participants, des modèles topographiques optimaux en utilisant les modèles de decoding linéaires décrit dans (Faghel-Soubeyrand et al., 2022). Ces modèles simples nous donneraient une topographie de poids EEG (e.g. 128 électrodes avec différents poids), qui servirait de cible à atteindre pendant dans un loop de *neurofeedback* contrôlé par apprentissage machine (Contestabile et al., 2002; Ordikhani-Seyedlar et al., 2016); pour une revue de l'utilisation du neurofeedback pour l'attention visuelle). De telles méthodes auraient des applications cliniques certaines en perception faciale, mais pourraient également être utilisées pour augmenter l'habileté d'individus avec des emplois requérant un traitement facial efficace.

Finalement, nos résultats empiriques fournissent des démonstrations importantes de la similarité entre les computations effectués au sein des modèles d'apprentissages profonds et les variations des fonctions humaines écologiques, i.e. le comportement perceptif humains. Des changements fonctionnels importants, comme par exemple la perte de l'habileté d'identification faciale, mènent à des changements de computations similaires à ceux des modèles d'apprentissages profonds visuels et langagiers dans le cerveau humain. Ces computations visuelles ou sémantiques de ces modèles

sont, oui, similaires au cerveau, mais, de façon plus importante, ne sont pas un épiphénomène puisqu'elles sont directement liées à des différences d'habileté en reconnaissances visuelles dans le monde réel.

7.4 Conclusion

En somme, en utilisant une approche diverse et multimodale et des populations d'individus diversifiées, cette thèse a révélé de nouveaux mécanismes perceptifs, cérébraux, et computationnels contribuant à notre habileté individuelle à reconnaître les visages. Ce faisant, elle participe à répondre à trois questions centrales des neurosciences cognitives et computationnelles : *quand* (i.e. le décours temporel cérébral), *quoi* (i.e. le contenu représentationnel) et *comment* (i.e. les computations cérébrales) le traitement des objets visuels de notre environnement affecte les idiosyncrasies de la perception et du comportement humain. De telles approches multimodales et computationnelles sont susceptibles de jouer un rôle essentiel dans le futur afin de révéler davantage la nature complexe de la perception, de la cognition et de la mémoire dans le cerveau humain.

Références bibliographiques

Achim, A. M., Ouellet, R., Lavoie, M.-A., Vallières, C., Jackson, P. L., & Roy, M.-A. (2013). Impact of social anxiety on social cognition and functioning in patients with recent-onset schizophrenia spectrum disorders. In *Schizophrenia Research* (Vol. 145, Issues 1-3, pp. 75–81). https://doi.org/10.1016/j.schres.2013.01.012

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72. https://doi.org/10.1038/nature03086

Adolphs, R., Spezio, M. L., Parlier, M., & Piven, J. (2008). Distinct face-processing strategies in parents of autistic children. *Current Biology: CB*, *18*(14), 1090–1093. https://doi.org/10.1016/j.cub.2008.06.073

Adrian, E. D., & Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *The Journal of Physiology*, *81*(4), 440–471. https://doi.org/10.1113/jphysiol.1934.sp003147

Alink, A., & Charest, I. (n.d.). *Individuals with clinically relevant autistic traits tend to have an eye for detail*. https://doi.org/10.1101/367532

Alink, A., & Charest, I. (2020). Clinically relevant autistic traits predict greater reliance on detail for image recognition. *Scientific Reports*, *10*(1), 14239. https://doi.org/10.1038/s41598-020-70953-8

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature*

Neuroscience, *25*(1), 116–126. https://doi.org/10.1038/s41593-021-00962-x

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., & Others. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126. https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/s41593-021-00962-x&casa_token=o9gTMfDJjm8AAAAA:1ZN_LewMpqfHZbsNx0LLta8DmmEbg90lftQSVycCcCXJvqhCHEvV7JLJcmonW7r2TWJjNVsHbBcY4iLIOXM

Alonso Prieto, E., Caharel, S., Henson, R., & Rossion, B. (2011). Early (N170/M170) Face-Sensitivity Despite Right Lateral Occipital Brain Damage in Acquired Prosopagnosia. *Frontiers in Human Neuroscience*, *5*, 138. https://doi.org/10.3389/fnhum.2011.00138

Ambrus, G. G., Windel, F., Mike Burton, A., & Kovács, G. (2017). Causal evidence of the involvement of the right occipital face area in face-identity acquisition. In *NeuroImage* (Vol. 148, pp. 212–218). https://doi.org/10.1016/j.neuroimage.2017.01.043

Andersen, S. K., & Hillyard, S. A. (2014). The time-course of feature-selective attention inside and outside the focus of spatial attention. In *Journal of Vision* (Vol. 14, Issue 10, pp. 23–23). https://doi.org/10.1167/14.10.23

Andersen, S. K., Hillyard, S. A., & Müller, M. M. (2008). Attention facilitates multiple stimulus features in parallel in human visual cortex. *Current Biology: CB*, *18*(13), 1006–1009. https://doi.org/10.1016/j.cub.2008.06.030

Andersen, S. K., Muller, M. M., & Hillyard, S. A. (2009). Color-selective attention need not be mediated by spatial attention. In *Journal of Vision* (Vol. 9, Issue 6, pp. 2–2). https://doi.org/10.1167/9.6.2

Anderson, A. J., McDermott, K., Rooks, B., Heffner, K. L., Dodell-Feder, D., & Lin, F. V. (2020). Decoding individual identity from brain activity elicited in imagining common experiences. *Nature Communications*, *11*(1), 5916. https://doi.org/10.1038/s41467-020-19630-y

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(3), 1128–1138. https://doi.org/10.1523/JNEUROSCI.4125-10.2011

Avidan, G., Hasson, U., Malach, R., & Behrmann, M. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 2. Functional Neuroimaging Findings. In *Journal of Cognitive Neuroscience* (Vol. 17, Issue 7, pp. 1150–1167). https://doi.org/10.1162/0898929054475145

Banville, H., Chehab, O., Hyvärinen, A., Engemann, D.-A., & Gramfort, A. (2021). Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, *18*(4). https://doi.org/10.1088/1741-2552/abca18

Behrmann, M., Avidan, G., Gao, F., & Black, S. (2007). Structural imaging reveals anatomical alterations in inferotemporal cortex in congenital prosopagnosia. *Cerebral Cortex* , *17*(10), 2354–2363. https://doi.org/10.1093/cercor/bhl144

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. https://doi.org/10.1162/jocn.1996.8.6.551

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller Primary Visual Cortex Is Associated with Stronger, but Less Precise Mental Imagery. *Cerebral Cortex* ,

26(9), 3838–3850. https://doi.org/10.1093/cercor/bhv186

Bobes, M. A., Lopera, F., Garcia, M., Díaz-Comas, L., Galan, L., & Valdes-Sosa, M. (2003). Covert matching of unfamiliar faces in a case of prosopagnosia: an ERP study. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *39*(1), 41–56. https://doi.org/10.1016/s0010-9452(08)70073-x

Bodamer, J. (1947). Die Prosop-Agnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, *179*(1), 6–53. https://doi.org/10.1007/BF00352849

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Brenner, C. A., Krishnan, G. P., Vohs, J. L., Ahn, W.-Y., Hetrick, W. P., Morzorati, S. L., & O'Donnell, B. F. (2009). Steady state responses: electrophysiological assessment of sensory function in schizophrenia. *Schizophrenia Bulletin*, *35*(6), 1065–1077. https://academic.oup.com/schizophreniabulletin/article-abstract/35/6/1065/1843589

Brockmole, J. R., & Logie, R. H. (2013). Age-Related Change in Visual Working Memory: A Study of 55,753 Participants Aged 8–75. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00012

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology* , *77 (Pt 3)*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bukach, C. M., Le Grand, R., Kaiser, M. D., Bub, D. N., & Tanaka, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology*, *2*(1), 227–244. https://doi.org/10.1348/174866407x231010

Bullier, J. (2001). Integrated model of visual processing. *Brain Research. Brain Research Reviews*, *36*(2-3), 96–107. https://doi.org/10.1016/s0165-0173(01)00085-6

Butler, P. D., & Javitt, D. C. (2005). Early-stage visual processing deficits in schizophrenia. *Current Opinion in Psychiatry*, *18*(2), 151–157. https://doi.org/10.1097/00001504-200503000-00008

Butler, P. D., Silverstein, S. M., & Dakin, S. C. (2008). Visual perception and its impairment in schizophrenia. *Biological Psychiatry*, *64*(1), 40–47. https://doi.org/10.1016/j.biopsych.2008.03.023

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex* , *6*(2), 81–92. https://doi.org/10.1093/cercor/6.2.81

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does Prosopagnosia Take the Eyes Out of Face Representations? Evidence for a Defect in Representing Diagnostic Facial Information following Brain Damage. In *Journal of Cognitive Neuroscience* (Vol. 17, Issue 10, pp. 1652–1666). https://doi.org/10.1162/089892905774597254

Caplette, L., Ince, R. A. A., Jerbi, K., & Gosselin, F. (2020). Disentangling presentation and processing times in the brain. *NeuroImage*, *218*. https://doi.org/10.1016/j.neuroimage.2020.116994

Caplette, L., Wicker, B., & Gosselin, F. (2016). Atypical Time Course of Object Recognition in Autism Spectrum Disorder. In *Scientific Reports* (Vol. 6, Issue 1). https://doi.org/10.1038/srep35494

Carlson, T. A., Grootswagers, T., & Robinson, A. K. (2019). An introduction to time-resolved decoding analysis for M/EEG. In *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/1905.04820

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1803.11175

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Charest, I., & Kriegeskorte, N. (2015). The brain of the beholder: honouring individual representational idiosyncrasies. *Language, Cognition and Neuroscience*, *30*(4), 367–379. https://doi.org/10.1080/23273798.2014.1002505

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. In *Journal of Vision* (Vol. 5, Issue 9, p. 1). https://doi.org/10.1167/5.9.1

Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J. F., & Charest, I. (2019). The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. In *NeuroImage* (Vol. 194, pp. 12–24). https://doi.org/10.1016/j.neuroimage.2019.03.031

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain

Responses in Space and Time. *Neuron*, *107*(5), 772–781. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Clark, C. M., Gosselin, F., & Goghari, V. M. (2013). Aberrant patterns of visual facial information usage in schizophrenia. *Journal of Abnormal Psychology*, *122*(2), 513–519. https://doi.org/10.1037/a0031944

Collette, C., Bonnotte, I., Jacquemont, C., Kalénine, S., & Bartolo, A. (2016). The Development of Object Function and Manipulation Knowledge: Evidence from a Semantic Priming Study. *Frontiers in Psychology*, *7*, 1239. https://doi.org/10.3389/fpsyg.2016.01239

Contestabile, M. T., Recupero, S. M., Palladino, D., De Stefanis, M., Abdolrahimzadeh, S., Suppressa, F., & Balacco Gabrieli, C. (2002). A new method of biofeedback in the management of low vision. *Eye* , *16*(4), 472–480. https://doi.org/10.1038/sj.eye.6700046

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: current perspectives. *Eye and Brain*, *8*, 165–175. https://doi.org/10.2147/EB.S92838

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review. *Schizophrenia Bulletin*, *32 Suppl 1*, S44–S63. https://doi.org/10.1093/schbul/sbl029

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J. S., Yonas, A., & Duchaine,

B. (2014). “A room full of strangers every day”: The psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research*, *77*(2), 144–150. https://doi.org/10.1016/j.jpsychores.2014.06.001

Dalrymple, K. A., Oruç, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., Handy, T. C., & Barton, J. J. S. (2011). The anatomic basis of the right face-selective N170 IN acquired prosopagnosia: a combined ERP/fMRI study. *Neuropsychologia*, *49*(9), 2553–2563. https://doi.org/10.1016/j.neuropsychologia.2011.05.003

Daube, C., Xu, T., Zhan, J., Webb, A., Ince, R. A. A., Garrod, O. G. B., & Schyns, P. G. (2021). Grounding deep neural network predictions of human categorization behavior in understandable functional features: The case of face identity. *Patterns (New York, N.Y.)*, *2*(10), 100348. https://doi.org/10.1016/j.patter.2021.100348

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, *23*(11), 1279–1287. https://doi.org/10.1177/0956797612446350

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), 1258. https://doi.org/10.1038/s41467-019-09239-1

Doerig, A., Kietzmann, T. C., Allen, E., Wu, Y., Naselaris, T., Kay, K., & Charest, I. (2022).

Semantic scene descriptions as an objective of human vision. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2209.11737

Doerig, A., Sommers, R., Seeliger, K., Richards, B., Ismael, J., Lindsay, G., Kording, K., Konkle, T., Van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2022). The neuroconnectionist research programme. In *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/2209.03718

Duan, Y., Zhan, J., Ince, R., van Rijsbergen, N., & Schyns, P. (2018). Transfer of Diagnostic Features from Occipital Cortex to right Fusiform Gyrus for Perceptual Decisions. In *Journal of Vision* (Vol. 18, Issue 10, p. 736). https://doi.org/10.1167/18.10.736

Dubois, J., & Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443. https://doi.org/10.1016/j.tics.2016.03.014

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*, 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, *17*(14), 7. https://doi.org/10.1167/17.14.7

Duncan, J., Royer, J., Dugas, G., Blais, C., & Fiset, D. (2019). Revisiting the link between horizontal tuning and face processing ability with independent measures. *Journal of Experimental Psychology. Human Perception and Performance*, *45*(11), 1429–1435. https://doi.org/10.1037/xhp0000684

Dunn, J. D., Varela, V. P. L., Nicholls, V. I., Papinutto, M., White, D., & Miellet, S. (2022).

Face-Information Sampling in Super-Recognizers. *Psychological Science*, *33*(9), 1615–1630. https://doi.org/10.1177/09567976221096320

Dwivedi, K., Bonner, M. F., Cichy, R. M., & Roig, G. (2021). Unveiling functions of the visual cortex using task-specific deep neural networks. *PLoS Computational Biology*, *17*(8), e1009267. https://doi.org/10.1371/journal.pcbi.1009267

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain: A Journal of Neurology*, *135*(Pt 2), 542–554. https://doi.org/10.1093/brain/awr347

Elbich, D. B., & Scherf, S. (2017). Beyond the FFA: Brain-behavior correspondences in face recognition abilities. *NeuroImage*, *147*, 409–422. https://doi.org/10.1016/j.neuroimage.2016.12.042

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125. https://doi.org/10.1016/s0896-6273(00)80758-8

Faghel-Soubeyrand, S., Alink, A., Bamps, E., Gervais, R.-M., Gosselin, F., & Charest, I. (2019). The two-faces of recognition ability: better face recognizers extract different physical content from left and right sides of face stimuli. *Journal of Vision*, *19*(10), 136d – 136d. https://doi.org/10.1167/19.10.136d

Faghel-Soubeyrand, S., Alink, A., Bamps, E., Gosselin, F., & Charest, I. (2019). Visual representations supporting category-specific information about visual objects in the brain. In *2019 Conference on Cognitive Computational Neuroscience*. https://doi.org/10.32470/ccn.2019.1404-0

Faghel-Soubeyrand, S., Dupuis-Roy, N., & Gosselin, F. (2019). Inducing the use of right eye enhances face-sex categorization performance. *Journal of Experimental Psychology. General*, *148*(10), 1834–1841. https://doi.org/10.1037/xge0000542

Faghel-Soubeyrand, S., & Gosselin, F. (2017). Task-modulated integration of facial features in the brain. *Journal of Vision*, *17*(10), 268. https://doi.org/10.1167/17.10.268

Faghel-Soubeyrand, S., Kloess, J. A., Gosselin, F., Charest, I., & Woodhams, J. (2021). Diagnostic Features for Human Categorisation of Adult and Child Faces. *Frontiers in Psychology*, *12*, 775338. https://doi.org/10.3389/fpsyg.2021.775338

Faghel-Soubeyrand, S., Lecomte, T., Bravo, M. A., Lepage, M., Potvin, S., Abdel-Baki, A., Villeneuve, M., & Gosselin, F. (2020). Abnormal visual representations associated with confusion of perceived facial expression in schizophrenia with social anxiety disorder. *NPJ Schizophrenia*, *6*(1), 28. https://doi.org/10.1038/s41537-020-00116-1

Faghel-Soubeyrand, S., Ramon, M., Bamps, E., Zoia, M., Woodhams, J., Richoz, A.-R., Caldara, R., Gosselin, F., & Charest, I. (2022). The neural code behind face recognition abilities. In *bioRxiv* (p. 2022.03.19.484245). https://doi.org/10.1101/2022.03.19.484245

Fiset, D., Blais, C., Royer, J., Richoz, A.-R., Dugas, G., & Caldara, R. (2017). Mapping the impairment in decoding static facial expressions of emotion in prosopagnosia. *Social Cognitive and Affective Neuroscience*, *12*(8), 1334–1341. https://doi.org/10.1093/scan/nsx068

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychological Medicine*, *39*(6), 889–905. https://doi.org/10.1017/S0033291708004558

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*(4), 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual Differences in Visual Working Memory Capacity. In *Mechanisms of Sensory Working Memory* (pp. 105–119). https://doi.org/10.1016/b978-0-12-801371-7.00009-0

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience*, *23*(7), 1723–1740. https://doi.org/10.1162/jocn.2010.21545

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233. https://doi.org/10.1098/rsos.200233

Gao, X., Vuong, Q. C., & Rossion, B. (2019). The cortical face network of the prosopagnosic patient PS with fast periodic stimulation in fMRI. In *Cortex* (Vol. 119, pp. 528–542). https://doi.org/10.1016/j.cortex.2018.11.008

Gauthier, I. (2018). Domain-Specific and Domain-General Individual Differences in Visual Object Recognition. *Current Directions in Psychological Science*, *27*(2), 97–102. https://doi.org/10.1177/0963721417737151

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

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. https://direct.mit.edu/jocn/article-abstract/12/3/495/3437

Gervais, R.-M., Faghel-Soubeyrand, S., Tardif, J., & Gosselin, F. (2021). Using EEG frequency-tagging to measure visual representations of faces. *Journal of Vision*, *21*(9), 2637–2637. https://doi.org/10.1167/jov.21.9.2637

Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *1*, 143. https://doi.org/10.3389/fpsyg.2010.00143

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*, 1–11. https://doi.org/10.1016/j.neuropsychologia.2015.12.004

Goffaux, V., & Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, *6*, 34204. https://doi.org/10.1038/srep34204

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://www.ncbi.nlm.nih.gov/pubmed/11448718

Graf, M., & Bülthoff, H. H. (2019). Object Shape in Basic Level Categorisation. In *Proceedings of EuroCogSci 03* (pp. 390–390). https://doi.org/10.4324/9781315782362-93

Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews. Neuroscience*, *16*(10), 620–631. https://doi.org/10.1038/nrn4005

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562.

https://doi.org/10.1038/nn1224

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10-11), 1409–1422. https://doi.org/10.1016/s0042-6989(01)00073-6

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews. Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1714). https://doi.org/10.1098/rstb.2016.0102

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex* , *20*(10), 2304–2318. https://doi.org/10.1093/cercor/bhp316

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. In *eLife* (Vol. 7). https://doi.org/10.7554/elife.32816

Heisz, J. J., Watter, S., & Shedden, J. M. (2006). Automatic face identity encoding at the N170. *Vision Research*, *46*(28), 4604–4614. https://doi.org/10.1016/j.visres.2006.09.026

Herzmann, G., Kunina, O., & Sommer, W. (2010). Individual differences in face cognition:

brain–behavior relationships. *Journal of Cognitive*. https://direct.mit.edu/jocn/article-abstract/22/3/571/4821

Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V., Villringer, A., & Gaebler, M. (2021). Decoding subjective emotional arousal from EEG during an immersive virtual reality experience. *eLife*, *10*. https://doi.org/10.7554/eLife.64812

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639–642. https://doi.org/10.1126/science.1234330

Huang, J., Chan, R. C. K., Gollan, J. K., Liu, W., Ma, Z., Li, Z., & Gong, Q.-Y. (2011). Perceptual bias of patients with schizophrenia in morphed facial expression. *Psychiatry Research*, *185*(1-2), 60–65. https://doi.org/10.1016/j.psychres.2010.05.017

Huang, L., Song, Y., Li, J., Zhen, Z., Yang, Z., & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Frontiers in Human Neuroscience*, *8*, 483. https://doi.org/10.3389/fnhum.2014.00483

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. In *Science* (Vol. 310, Issue 5749, pp. 863–866).

https://doi.org/10.1126/science.1117593

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573. https://doi.org/10.1002/hbm.23471

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. In *Cerebral Cortex* (Vol. 26, Issue 11, pp. 4123–4135). https://doi.org/10.1093/cercor/bhw196

Issa, E. B., & DiCarlo, J. J. (2012). Precedence of the eye region in neural processing of faces. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(47), 16666–16682. https://doi.org/10.1523/JNEUROSCI.2391-12.2012

Jacques, C., Schiltz, C., & Goffaux, V. (2014). Face perception is tuned to horizontal orientation in the N170 time window. *Journal of Vision*, *14*(2). https://doi.org/10.1167/14.2.5

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), E6418–E6427. https://doi.org/10.1073/pnas.1802246115

Jonas, J., Rossion, B., Brissart, H., Frismand, S., Jacques, C., Hossu, G., Colnat-Coulbois, S., Vespignani, H., Vignal, J.-P., & Maillard, L. (2015). Beyond the core face-processing network: Intracerebral stimulation of a face-selective area in the right anterior fusiform

gyrus elicits transient prosopagnosia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *72*, 140–155. https://doi.org/10.1016/j.cortex.2015.05.026

Jonas, J., Rossion, B., Krieg, J., Koessler, L., Colnat-Coulbois, S., Vespignani, H., Jacques, C., Vignal, J.-P., Brissart, H., & Maillard, L. (2014). Intracerebral electrical stimulation of a face-selective area in the right inferior occipital cortex impairs individual face discrimination. *NeuroImage*, *99*, 487–497. https://doi.org/10.1016/j.neuroimage.2014.06.017

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1187–1211. https://doi.org/10.1152/jn.1987.58.6.1187

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, *83*, 201–226. https://doi.org/10.1016/j.neuropsychologia.2015.10.023

Jozwik, K. M., Najarro, E., van den Bosch, J. J. F., Charest, I., Cichy, R. M., & Kriegeskorte, N. (2022). Disentangling five dimensions of animacy in human brain and behaviour. *Communications Biology*, *5*(1), 1247. https://doi.org/10.1038/s42003-022-04194-y

Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 861–878. https://doi.org/10.3758/s13415-013-0234-y

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8),

759–763. https://doi.org/10.1038/77664

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311. https://www.ncbi.nlm.nih.gov/pubmed/9151747

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713

Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *eLife*, *6*. https://doi.org/10.7554/eLife.22341

Khaligh-Razavi, S. M., Bainbridge, W. A., & Pantazis, D. (2016). From what we perceive to what we remember: Characterizing representational dynamics of visual memorability. *bioRxiv*. https://www.biorxiv.org/content/10.1101/049700v1.abstract

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kim, D.-W., Shim, M., Song, M. J., Im, C.-H., & Lee, S.-H. (2015). Early visual processing deficits in patients with schizophrenia during spatial frequency-dependent facial affect processing. *Schizophrenia Research*, *161*(2-3), 314–321. https://doi.org/10.1016/j.schres.2014.12.020

Kogata, T., & Iidaka, T. (2018). A review of impaired visual processing and the daily visual world in patients with schizophrenia. *Nagoya Journal of Medical Science*, *80*(3), 317–328. https://doi.org/10.18999/nagjms.80.3.317

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, *21*(1), 38–43. https://doi.org/10.1177/0956797609356508

Kragel, P. A., & LaBar, K. S. (2016). Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Sciences*, *20*(6), 444–455. https://doi.org/10.1016/j.tics.2016.03.011

Kriegeskorte, N. (2015a). Deep neural networks: a new framework for modelling biological vision and brain information processing. In *bioRxiv* (p. 029876). https://doi.org/10.1101/029876

Kriegeskorte, N. (2015b). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*, *1*, 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

Kriegeskorte, N., & Diedrichsen, J. (2016). Inferring brain-computational mechanisms with models of activity measurements. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1705). https://doi.org/10.1098/rstb.2016.0278

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., & Wei, X.-X. (2021). Neural tuning and representational geometry. *Nature Reviews. Neuroscience*, *22*(11), 703–718. https://doi.org/10.1038/s41583-021-00502-3

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/s1364-6613(00)01560-6

Langner, O., Becker, E. S., & Rinck, M. (2009). Social anxiety and anger identification: bubbles reveal differential use of facial information with low spatial frequencies. *Psychological Science*, *20*(6), 666–670. https://doi.org/10.1111/j.1467-9280.2009.02357.x

Lecomte, T., Théroux, L., Paquin, K., Potvin, S., & Achim, A. (2019). Can Social Anxiety Impact Facial Emotion Recognition in Schizophrenia? *The Journal of Nervous and Mental Disease*, *207*(3), 140–144. https://doi.org/10.1097/NMD.0000000000000934

Lee, J., & Geng, J. J. (2017). Idiosyncratic Patterns of Representational Similarity in Prefrontal Cortex Predict Attentional Performance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(5), 1257–1268. https://doi.org/10.1523/JNEUROSCI.1407-16.2016

Lee, J., Gosselin, F., Wynn, J. K., & Green, M. F. (2011). How do schizophrenia patients use visual information to decode facial emotion? *Schizophrenia Bulletin*, *37*(5), 1001–1008. https://doi.org/10.1093/schbul/sbq006

Levinson, M., Podvalny, E., Baete, S. H., & He, B. J. (2021). Cortical and subcortical signatures of conscious object recognition. *Nature Communications*, *12*(1), 2930. https://doi.org/10.1038/s41467-021-23266-x

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M. D., Crutch, S., Woodward, F., Leung, K., Fox, N. C., & Husain, M. (2016). Visual short-term memory binding deficit in familial Alzheimer's disease. In *Cortex* (Vol. 78, pp. 150–164). https://doi.org/10.1016/j.cortex.2016.01.015

Lindh, D., Sligte, I. G., Assecondi, S., Shapiro, K. L., & Charest, I. (2019). Conscious perception of natural images is constrained by category-related visual features. In *Nature Communications* (Vol. 10, Issue 1). https://doi.org/10.1038/s41467-019-12135-3

Li, S. P. D., & Bonner, M. F. (2021). Tuning in scene-preferring cortex for mid-level visual features gives rise to selectivity across multiple levels of stimulus complexity. In *bioRxiv* (p. 2021.09.24.461733). https://doi.org/10.1101/2021.09.24.461733

Liu-Shuang, J., Torfs, K., & Rossion, B. (2016). An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia*, *83*, 100–113. https://doi.org/10.1016/j.neuropsychologia.2015.08.023

Lohse, M., Garrido, L., Driver, J., Dolan, R. J., Duchaine, B. C., & Furl, N. (2016). Effective Connectivity from Early Visual Cortex to Posterior Occipitotemporal Face Areas Supports

Face Selectivity and Predicts Developmental Prosopagnosia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(13), 3821–3828. https://doi.org/10.1523/JNEUROSCI.3621-15.2016

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38), E9015–E9024. https://doi.org/10.1073/pnas.1719616115

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. In *Electroencephalography and Clinical Neurophysiology* (Vol. 75, Issue 6, pp. 528–542). https://doi.org/10.1016/0013-4694(90)90139-b

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/s1364-6613(02)01903-4

McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *12*(1), 76–82. https://doi.org/10.1016/s0010-9452(76)80033-0

Minnebusch, D. A., Suchan, B., Köster, O., & Daum, I. (2009). A bilateral occipitotemporal network mediates face perception. *Behavioural Brain Research*, *198*(1), 179–185. https://doi.org/10.1016/j.bbr.2008.10.041

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(10), 4770–4774. https://doi.org/10.1073/pnas.93.10.4770

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555–604. https://doi.org/10.1162/jocn.1997.9.5.555

Naselaris, T., Allen, E., & Kay, K. (2021). Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences*, *40*, 45–51. https://doi.org/10.1016/j.cobeha.2020.12.008

Naselaris, T., Bassett, D. S., Fletcher, A. K., Kording, K., Kriegeskorte, N., Nienborg, H., Poldrack, R. A., Shohamy, D., & Kay, K. (2018). Cognitive Computational Neuroscience: A New Conference for an Emerging Discipline. *Trends in Cognitive Sciences*, *22*(5), 365–367. https://doi.org/10.1016/j.tics.2018.02.008

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes.

NeuroImage, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Nonaka, S., Majima, K., Aoki, S. C., & Kamitani, Y. (2021). Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *iScience*, *24*(9), 103013. https://doi.org/10.1016/j.isci.2021.103013

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Norton, D., McBain, R., Holt, D. J., Ongur, D., & Chen, Y. (2009). Association of impaired facial affect recognition with basic facial and visual processing deficits in schizophrenia. *Biological Psychiatry*, *65*(12), 1094–1098. https://doi.org/10.1016/j.biopsych.2009.01.026

Nowparast Rostami, H., Sommer, W., Zhou, C., Wilhelm, O., & Hildebrandt, A. (2017). Structural encoding processes contribute to individual differences in face and object cognition: Inferences from psychometric test performance and event-related brain potentials. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *95*, 192–210. https://doi.org/10.1016/j.cortex.2017.08.017

Ordikhani-Seyedlar, M., Lebedev, M. A., Sorensen, H. B. D., & Puthusserypady, S. (2016). Neurofeedback Therapy for Enhancing Visual Attention: State-of-the-Art and Challenges. *Frontiers in Neuroscience*, *10*, 352. https://doi.org/10.3389/fnins.2016.00352

Oruc, I., Shafai, F., & Iarocci, G. (2018). Link Between Facial Identity and Expression Abilities Suggestive of Origins of Face Impairments in Autism: Support for the Social-Motivation Hypothesis. *Psychological Science*, *29*(11), 1859–1867. https://doi.org/10.1177/0956797618795471

Pachai, M. V., Sekuler, A. B., & Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, *4*, 74. https://doi.org/10.3389/fpsyg.2013.00074

Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in Brain Research*, *165*, 493–507. https://doi.org/10.1016/S0079-6123(06)65031-0

Parkkonen, L., Andersson, J., Hämäläinen, M., & Hari, R. (2008). Early visual brain areas reflect the percept of an ambiguous scene. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20500–20504. https://doi.org/10.1073/pnas.0810966105

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(24), 6171–6176. https://doi.org/10.1073/pnas.1721355115

Popal, H., Wang, Y., & Olson, I. R. (2019). A Guide to Representational Similarity Analysis for Social Neuroscience. *Social Cognitive and Affective Neuroscience*, *14*(11), 1243–1253. https://doi.org/10.1093/scan/nsz099

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. In *Nature Neuroscience* (Vol. 24, Issue 11, pp. 1628–1636). https://doi.org/10.1038/s41593-021-00921-6

Premkumar, P., Cooke, M. A., Fannon, D., Peters, E., Michel, T. M., Aasen, I., Murray, R. M.,

Kuipers, E., & Kumari, V. (2008). Misattribution bias of threat-related facial expressions is related to a longer duration of illness and poor executive function in schizophrenia and schizoaffective disorder. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *23*(1), 14–19. https://doi.org/10.1016/j.eurpsy.2007.10.004

Proix, T., Delgado Saa, J., Christen, A., Martin, S., Pasley, B. N., Knight, R. T., Tian, X., Poeppel, D., Doyle, W. K., Devinsky, O., Arnal, L. H., Mégevand, P., & Giraud, A.-L. (2022). Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nature Communications*, *13*(1), 48. https://doi.org/10.1038/s41467-021-27725-3

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, *15*(9), 15. https://doi.org/10.1167/15.9.15

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281–1302. https://doi.org/10.1037/a0037004

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews. Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Rossion, B. (2022a). Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part I: Function. In *Neuropsychologia* (Vol. 173, p. 108278). https://doi.org/10.1016/j.neuropsychologia.2022.108278

Rossion, B. (2022b). Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part II: Neural basis. *Neuropsychologia*, 108279. https://doi.org/10.1016/j.neuropsychologia.2022.108279

Rossion, B., Caldara, R., Seghier, M., Schuller, A., Lazeyras, F., & Mayer, E. (2003). A network of occipito temporal face sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain: A Journal of Neurology*, *126*(11), 2381–2395. https://doi.org/10.1093/brain/awg241

Rossion, B., & Jacques, C. (2012). The N170: Understanding the time course of face perception in the human brain. *The Oxford Handbook of Event-Related Potential Components.*, *641*, 115–141. https://psycnet.apa.org/fulltext/2013-01016-005.pdf

Rossion, B., Kaiser, M. D., Bub, D., & Tanaka, J. W. (2009). Is the loss of diagnosticity of the eye region of the face a common aspect of acquired prosopagnosia? *Journal of Neuropsychology*, *3*(Pt 1), 69–78. https://doi.org/10.1348/174866408X289944

Rossion, B., Retter, T. L., & Liu-Shuang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *The European Journal of Neuroscience*, *52*(10), 4283–4344. https://doi.org/10.1111/ejn.14865

Roth, Z. N., Kay, K., & Merriam, E. P. (2022). Natural scene sampling reveals reliable coarse-scale orientation tuning in human V1. *Nature Communications*, *13*(1), 6469.

https://doi.org/10.1038/s41467-022-34134-7

Rousselet, G. A., Ince, R. A. A., van Rijsbergen, N. J., & Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, *14*(13), 7. https://doi.org/10.1167/14.13.7

Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. In *Cognition* (Vol. 181, pp. 12–20). https://doi.org/10.1016/j.cognition.2018.08.004

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Rust, N. C., & DiCarlo, J. J. (2012). Balanced Increases in Selectivity and Tolerance Produce Constant Sparseness along the Ventral Visual Stream. In *Journal of Neuroscience* (Vol. 32, Issue 30, pp. 10170–10182). https://doi.org/10.1523/jneurosci.6125-11.2012

Saari, P., Burunat, I., Brattico, E., & Toiviainen, P. (2018). Decoding Musical Training from Dynamic Processing of Musical Features in the Brain. *Scientific Reports*, *8*(1), 708. https://doi.org/10.1038/s41598-018-19177-5

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, *79*(6), 3272–3278. https://doi.org/10.1152/jn.1998.79.6.3272

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. https://doi.org/10.1038/nn.2706

Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *Neuroreport*, *14*(13), 1665–1669. https://doi.org/10.1097/00001756-200309150-00002

Schyns, P. G., Snoek, L., & Daube, C. (2022). Degrees of algorithmic equivalence between the brain and its DNN models. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.09.003

Schyns, P. G., Zhan, J., Jack, R. E., & Ince, R. A. A. (2020). Revealing the information contents of memory within the stimulus information representation framework. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1799), 20190705. https://doi.org/10.1098/rstb.2019.0705

Shen, C., Stasch, J., Velenosi, L., Madipakkam, A. R., Edemann-Callesen, H., & Neuhaus, A. H. (2017). Face identity is encoded in the duration of N170 adaptation. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *86*, 55–63. https://doi.org/10.1016/j.cortex.2016.10.010

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., Boly, M., Postle, B. R., & Tononi, G. (2017). The neural correlates of dreaming. *Nature Neuroscience*, *20*(6), 872–878. https://doi.org/10.1038/nn.4545

Sigurdardottir, H. M., & Gauthier, I. (2015). Expertise and Object Recognition. In *Brain Mapping* (pp. 523–527). https://doi.org/10.1016/b978-0-12-397025-1.00038-5

Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., Gosselin, F., Hadley, D. M., Cavanagh, J., & Schyns, P. G. (2008). Classification images reveal the information sensitivity of brain voxels in fMRI. *NeuroImage*, *40*(4), 1643–1654.

https://doi.org/10.1016/j.neuroimage.2008.01.029

Smith, M. L., Gosselin, F., Cottrell, G. W., & Schyns, P. G. (2004). Transmitting and decoding facial expressions of emotion. In *Journal of Vision* (Vol. 4, Issue 8, pp. 909–909). https://doi.org/10.1167/4.8.909

Smith, M. L., Gosselin, F., & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science*, *15*(11), 753–761. https://doi.org/10.1111/j.0956-7976.2004.00752.x

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring Internal Representations from Behavioral and Brain Data. In *Current Biology* (Vol. 22, Issue 3, pp. 191–196). https://doi.org/10.1016/j.cub.2011.11.061

Smith, M. L., Lestou, V., Gosselin, F., & Schyns, P. G. (2009). Visualizing Internal Representations from Behavioral and Brain Imaging Data. In *NeuroImage* (Vol. 47, p. S63). https://doi.org/10.1016/s1053-8119(09)70308-8

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *NeuroImage*, *35*(2), 836–852. https://doi.org/10.1016/j.neuroimage.2006.09.051

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*(5), 929–939. https://doi.org/10.1007/s10803-006-0232-9

Stein, T., Reeder, R. R., & Peelen, M. V. (2016). Privileged access to awareness for faces and objects of expertise. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 42, Issue 6, pp. 788–798). https://doi.org/10.1037/xhp0000188

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. In *Current Opinion in Neurobiology* (Vol. 23, Issue 3, pp. 423–429). https://doi.org/10.1016/j.conb.2012.12.011

Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic face processing. *The Oxford Handbook of Face Perception*, 177–194. https://books.google.com/books?hl=en&lr=&id=HunrfhMRx_8C&oi=fnd&pg=PA177&dq=holistic+face+processing+review&ots=OmjorJ-VF6&sig=dZM8kQbFJB47oMkCSjGZdNL-R-0

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://www.sciencedirect.com/science/article/pii/001002859190016H

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*(5134), 685–688. https://doi.org/10.1126/science.8235589

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019a). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science*, *30*(2), 300–308. https://journals.sagepub.com/doi/abs/10.1177/0956797618811338

Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019b). Use of Face Information Varies Systematically From Developmental Prosopagnosics to Super-Recognizers. *Psychological Science*, *30*(2), 300–308. https://doi.org/10.1177/0956797618811338

Taubert, J., Goffaux, V., Van Belle, G., Vanduffel, W., & Vogels, R. (2016). The impact of

orientation filtering on face-selective neurons in monkey inferior temporal cortex. *Scientific Reports*, *6*, 21189. https://doi.org/10.1038/srep21189

Thompson, J. A. F. (2021). Forms of explanation and understanding for neuroscience and artificial intelligence. *Journal of Neurophysiology*, *126*(6), 1860–1874. https://doi.org/10.1152/jn.00195.2021

Towler, A., Keshwa, M., Ton, B., Kemp, R. I., & White, D. (2021). Diagnostic feature training improves face matching accuracy. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *47*(8), 1288–1298. https://doi.org/10.1037/xlm0000972

Towler, J., & Eimer, M. (2012). Electrophysiological studies of face processing in developmental prosopagnosia: neuropsychological and neurodevelopmental perspectives. *Cognitive Neuropsychology*, *29*(5-6), 503–529. https://doi.org/10.1080/02643294.2012.716757

Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670–674. https://doi.org/10.1126/science.1119983

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. In *Nature Neuroscience* (Vol. 5, Issue 7, pp. 682–687). https://doi.org/10.1038/nn870

Vinken, K., Konkle, T., & Livingstone, M. (2022). The neural code for “face cells” is not face specific. In *bioRxiv* (p. 2022.03.06.483186). https://doi.org/10.1101/2022.03.06.483186

Visconti di Oleggio Castello, M., Haxby, J. V., & Gobbini, M. I. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(45). https://doi.org/10.1073/pnas.2110474118

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Weisberg, J., van Turennout, M., & Martin, A. (2007). A neural system for learning about object function. *Cerebral Cortex* , *17*(3), 513–521. https://doi.org/10.1093/cercor/bhj176

White, D., & Mike Burton, A. (2022). Individual differences and the multidimensional nature of face perception. In *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-022-00041-3

Williams, M. A., Berberovic, N., & Mattingley, J. B. (2007). Abnormal FMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Current Biology: CB*, *17*(14), 1259–1264. https://doi.org/10.1016/j.cub.2007.06.042

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*, *29*(5-6), 360–392.

https://doi.org/10.1080/02643294.2012.753433

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

Zhang, J., Liu, J., & Xu, Y. (2015). Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(4), 1539–1548. https://doi.org/10.1523/JNEUROSCI.2646-14.2015

Zhan, J., Ince, R. A. A., van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic Construction of Reduced Representations in the Brain for Perceptual Decision Behavior. *Current Biology: CB*, *29*(2), 319–326.e4. https://doi.org/10.1016/j.cub.2018.11.049

Annexes

Annexe I : Article 6

Visual representations supporting category-specific information about visual objects in the brain

Faghel-Soubeyrand[1,2], S., Alink[3], A., Bamps, E[4]., Gosselin, F*[1]. & Charest, I*[1,2].

1. Département de psychologie, Université de Montréal, Canada
2. Centre for Human Brain Health, University of Birmingham, UK
3. University of Hamburg, Faculty of Psychology and Human Movement, Germany
4. Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven,

January 2019

DOI:10.32470/CCN.2019.1404-0

Pulished (Conference paper). Faghel-Soubeyrand, S., Alink, A., Bamps, E., Gosselin, F. and Charest, I. (2019). Visual representations supporting category-specific information about visual objects in the brain. Cognitive Computational Neuroscience, Berlin. Conference paper.

Abstract

Over recent years, multivariate pattern analysis ("decoding") approaches have become increasingly used to investigate "when" and "where" our brains conduct meaningful processes about their visual environments. Studies using time-resolved decoding of M/EEG patterns have described numerous processes such as object/face familiarity and the emergence of basic-to-abstract category information. Surprisingly, no study has, to our knowledge, revealed "what" (i.e. the actual visual information that) our brain uses while these computations are examined by decoding algorithms. Here, we revealed the time course at which our brain extracts realistic category-specific information about visual objects (i.e. emotion-type & gender information from faces) with time-resolved decoding of high-density EEG patterns, as well as carefully controlled tasks and visual stimulation. Then, we derived temporal generalization matrices and showed that category-specific information is 1) first diffused across brain areas (250 to 350 ms) and 2) encoded under a stable neural pattern that suggest evidence accumulation (350 to 650 ms after face onset). Finally, we bridged time-resolved decoding with psychophysics and revealed the specific visual information (spatial frequency, feature position & orientation information) that support these brain computations. Doing so, we uncovered interconnected dynamics between visual features, and the accumulation and diffusion of category-specific information in the brain.

Keywords: vision; categorical representation; EEG; multivariate pattern analysis; psychophysics

Introduction

Recognizing a visual object requires the timely processing of visual information, from fine grained low-level features, to more abstract category-relevant information. In the recent years, a wealth of object recognition neuroimaging results has been obtained using multivariate pattern analyses ("decoding") to characterize "when" and "where" our brains process information about specific stimuli and tasks. To resolve the temporal dynamics of object recognition, studies have applied linear classifiers to time-resolved Magneto/electro-encephalography (M/EEG) activity patterns (see Carlson et al. 2019 for a review). This decoding approach yielded important understandings in objects/face familiarity (Dobs et al. 2019), object memorability (Mohsenzadeh et al. 2019), the emergence of basic-to-abstract category information (Cichy et al. 2014; Contini et al. 2017) to name only but a few studies.

To form a richer understanding of object categorization in the brain at a mechanistic level, such decoding approaches need to be combined with psychophysical procedures that enable revealing the specific content of the information that is available to the brain during a cognitive task. Along this line, a recent study using MEG and psychophysics revealed the processing of task-relevant information in the brain (Zhan et al. 2019),

with increasing importance of behaviourally relevant information along the visual ventral stream. These results culminated from over a decade of work using psychophysical techniques similar to reverse correlation (e.g. see Bubbles, Gosselin and Schyns 2001) with behavioral or brain imaging data to decipher the specific visual information (e.g. "detailed" vs. "coarse" spatial frequency information) underlying object (Caplette et al., 2016), scene (Willenbockel, Gosselin & Võ, submitted) and face (Smith et al. 2008) recognition.

Combining tightly controlled psychophysical paradigms with decoding algorithms would not only provide a way to investigate categorical representations while characterizing the visual features that support information processing in the brain. If used properly, it would also enable neuroscientists to control for low-level processing interpretations behind the disclosed computations. Indeed, it can be a complex endeavour to untangle computations from low-level areas such as V1, which are sensitive to contrast and edge differences between face images from different gender, from decoded whole-brain information. Ultimately, even output from the retina could dissociate between a human face and a giraffe's; the goal of a neuroscientist studying high-level cognition, thus, is to describe how or when *category-specific* (e.g. human vs. animal) information is implemented in the brain.

Here, in an effort to bridge the gap between decoding studies and reverse correlation approaches, we combined diagnostic feature mapping —a carefully controlled psychophysical paradigm that reveals the features supporting recognition (Alink and Charest 2018)— with multivariate pattern analyses applied to concurrently measured EEG data. Similar to classical decoding approaches, we first reveal the time course at which our brain extracts realistic category-specific information about visual objects (i.e. emotion-type & gender information from faces). Then, we go a step further by revealing what visual information (spatial frequency, feature position & orientation information) supports category-specific computations in the brain.

Method

Experimental procedure & stimuli

Participants (N=5) were asked to categorize the emotion (fear vs. joy) or gender (male vs. female) of faces which physical content were partially revealed on every trial (2560 trials per task, per participant). More specifically, we presented 16 expressive male/female faces that varied in feature position content (e.g. more left-eye information present, no mouth information present), spatial-frequency content (coarseness of the presented information) and orientation content (e.g. more horizontal content) in a pseudo-random fashion on every trial (Figure 1; see Alink & Charest, 2018). This ensured that computations from low-level brain areas sensitive to these information variations (e.g. striate cortex), could not help our decoding algorithms, and, conversely, that higher-level information would be distilled-out. Both time-resolved brain information about emotion-type and gender were obtained from brains receiving this pseudo-random low-level visual

stimulation, and the original 16 face images used for this stimulation were identical between tasks. Thus, only task instructions ("discriminate emotion", "discriminate gender") differed.

A typical trial consisted of the following events : a fixation dot (with a jittered duration centered around .5s), and a randomly sampled face (1s) followed by the participant's response. Face images were presented in a random order across trials. The two task-conditions were completed on two separate EEG recordings and lasted ~3 hours each, including EEG headset installation.

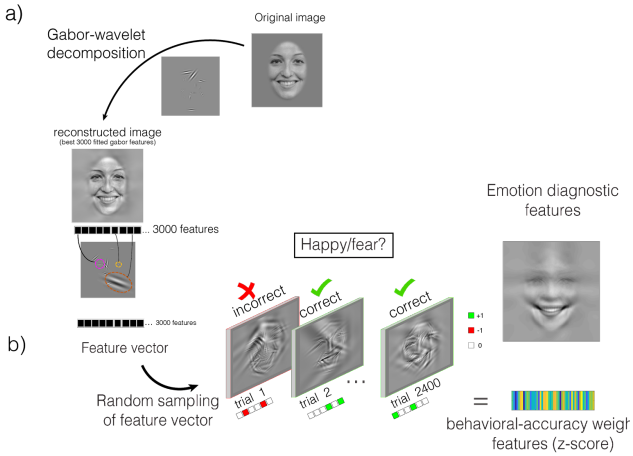


Figure 8.1. Stimuli creation and experimental paradigm. (a) A face image is fitted with Gabor patches of different orientations (10 to 180 deg), spatial-frequencies (from 2 to 76 cpi), [x,y] image coordinates and sizes. Only the 3k Gabors with the best fits (i.e. the features) are chosen to reconstruct an efficient version of the face image. (b) On a given trial, the feature vector is randomly sampled to create a stimulus.

Results

EEG recording & time resolved decoding

Electroencephalographic (EEG) data were recorded with a BioSemi 128 electrodes headset at 1024 hz. EEG traces were down sampled to 256 Hz and band-passed filtered (.01-30hz). For both task-conditions, the raw EEG patterns were fed to a linear-discriminant classifier which task was to classify the category of the presented stimulus (i.e. either gender in the face gender discrimination task, emotion-type in the emotion face discrimination task). We trained and tested (5-fold cross-validation, 5 repetitions) a different model with EEG patterns from every time point from -100 ms to 1000 ms in 4 ms steps, relative to face-onset. To assess significance, we trained and tested a classifier with identical parameters to categorize shuffled labels, repeated this step 10000 times, and compared our observed classification accuracy to this null-distribution in a time-resolved manner.

These decoding results are presented in **Figure 2a**. Specifically, we display the time course of this category-specific information for emotion-type (while participants completed the face-emotion discrimination task). It shows that emotion-specific information extracted by the brain from faces is present relatively late, starting from 272 ms ($p<.01$ points are underlined in gray; $p<.001$ in black), and culminates in a large temporal plateau around 550 ms. Furthermore, we observed a second informative window around 900ms after face onset. Traditional univariate event-related potentials (electrodes P6 & P7 average) are overlaid on top of the classification time-course. A brief comparison suggests that the decoded information —which discarded most of low-level information processes due to our controlled visual stimulation— emerges right

after face (N170) and attention-related processes (P200).

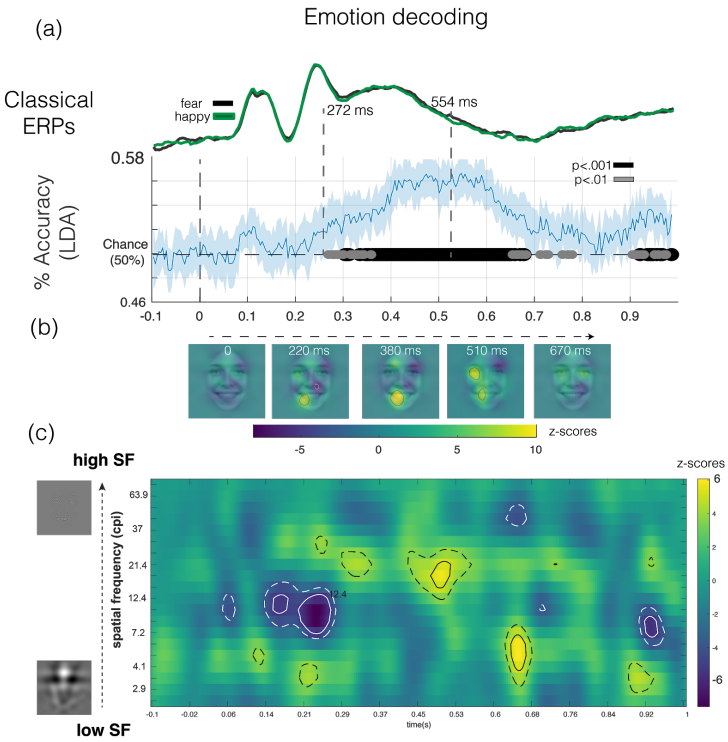


Figure 8.2. (a) face category-specific information extracted by the brain is unraveled in time using a linear-classifier. **(b, c)** The specific (feature position, spatial-frequency) visual representations supporting the decoded information is obtained by correlating the classifier's confidence-values with the presented features in a time resolved manner.

Temporal generalization

We then asked whether the temporal unfolding of this category-specific information emerged from a stable or varying neural code, i.e. if the EEG pattern dissociating two classes generalized at different points in time (King and Dehaene 2014). We computed a temporal generalization matrix where our lda model was trained on EEG patterns from a specific time *i* and tested on EEG patterns from a different time, *j*, within the -100 to 1000 ms time window. Accurate decoding at a specific coordinate in this matrix indicates that the model was able to generalize it's training at time *i* to time *j*, and therefore that both time points shared a similar (stable) EEG pattern.

The resulting matrix, shown in **figure 3**, indicates that our time-resolved decoding results emerged from at least two distinct brain processes. First, emotion category-specific information travels across brain areas from ~270 to 350 ms after face onset. This is then followed by a stable neural pattern that suggests evidence accumulation from ~350 to 650 ms. The following diagonal pattern suggests that emotion-category specific is finally diffused across brain areas, presumably before the perceptual decision.

Visual features behind category-specific information

But what, exactly, is contained in this category-specific information? To reveal the specific visual features underlying time resolved decoded information in the brain, we first extracted the trial-by-trial distance to the linear discriminant classifier decision criterion *c*, for every time point, participant and tasks. This produced a vector of 2560 distance-values (henceforth referred as *d-vals*) per participant, task and time point. These d-vals could be negative (class 1, e.g. happy category) or positive (class 2, e.g. fear category). To ensure that a positive correlation between visual features and d-vals

relate to accurate model classification, we sign-flipped (*-1) the d-vals for trials where class 1 stimuli were presented, such that any positive valued d-val reflects correct identification by the model, and any negative valued d-val reflects misclassification by the model. D-vals distributions were then z-transformed across the 2560 trials for every time point, participant, and task, and smoothed in the time domain with a Gaussian kernel of 4 ms of standard-deviation. Next, we computed the relative presence of the visual features across trials, for either spatial frequency content (from 2.4 cpi to 76.7 cpi in 20 steps), orientation content (10 to 180 degree in 10 degree steps) or feature [x,y] position (3000 image coordinates). As a final final step, we weighted this standardized [FeaturePresence x Trials] matrix with the standardized model confidence-values at every time point, and smoothed this matrix with a 2D Gaussian of 2 standard-deviation. Again, permutation testing was used to assess significance.

This procedure revealed how brain representation of emotion categorical information is supported by spatial [x,y] feature position representations (**Figure 2b**). **Figure 2c** shows how this time-resolved information is supported by spatial-frequency (SF, coarse or fine details) content through time. Positive (yellow-green) values indicates a positive correlation between (accurate) model classification confidence and feature information presence, while a negative (dark-blue) values indicates low-confidence (uncertainty) of the model while the specific visual content was presented. In other words, high positive values indicates that a specific visual feature helped/supported the brain's representation of category-specific information from a face.

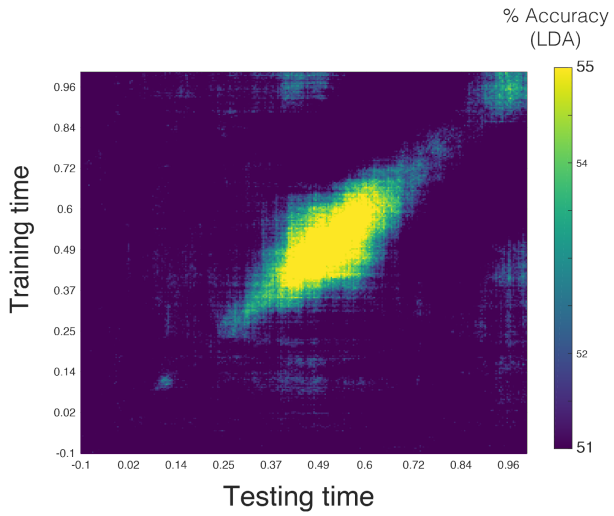


Figure 8.3. Generalization of EEG patterns through time. Above chance accuracy on the diagonal indicates variable brain representations, while off-diagonal decoding indicates a stable neural encoding of category-specific information.

A number of things can be distilled from the planes shown in figure 2 b and c. First, the peak decoding plateau observed from time-resolved classification (around 550 ms) is helped by the presence of mid-high-frequency, fine detailed visual information (~12-35cpi). Note that this 300-500 ms time window also coincides with the stable accumulation of evidence found in the temporal generalization matrix (figure 3). Second, a coarse-fine-coarse pattern seems to coincide with the diffusion-accumulation-diffusion of information that was shown in figure 3. Third, and finally, feature position maps indicates that high-level information about emotion is supported by clearly defined representation of facial attributes: mouth

representation supporting decoding in the 200-400 ms window, followed by a more complete mouth + left-eye representation coinciding with peak decoding accuracy.

Conclusion

Past studies have focused on describing where perceptual and cognitive representations were encoded or when they unfolded in time, but had yet to explicitly describe, and actually *see* the specific visual content that supports such representations. Here, we fill the gap between time-resolved decoding and visual psychophysics and reveal the visual features underlying the decoding of realistic, category-specific information in the brain through time. The visual features we reveal with this method match psychophysical behavioral data (e.g. coarse-to-fine processes, e.g. Caplette et al., 2016; mouth for emotion, Smith et al., 2008). Doing so, we show that seemingly simple linear steps of evidence processing by the visual system in fact engage multiple and interconnected dynamics between visual features, and the accumulation and diffusion of category-specific information in the brain.

Acknowledgments

This work was supported by an European Research Council (ERC) Starting Grant ERC-2017-StG 759432 (to I.C.), and a Natural Sciences and Engineering Research Council of Canada (NSERC) scholarship (to S.F.S).

References

Alink, Arjen, and Ian Charest. n.d. “Individuals with Clinically Relevant Autistic Traits Tend to Have an Eye for Detail.” https://doi.org/10.1101/367532.

Caplette, Laurent, Bruno Wicker, and Frédéric Gosselin. 2016. “Atypical Time Course of Object Recognition in Autism Spectrum Disorder.” *Scientific Reports*. https://doi.org/10.1038/srep35494.

Cichy, Radoslaw Martin, Dimitrios Pantazis, and Aude Oliva. 2014. “Resolving Human Object Recognition in Space and Time.” *Nature Neuroscience* 17 (3): 455–62.

Contini, Erika W., Susan G. Wardle, and Thomas A. Carlson. 2017. “Decoding the Time-Course of Object Recognition in the Human Brain: From Visual Features to Categorical Decisions.” *Neuropsychologia* 105 (October): 165–76.

Gosselin, F., and P. G. Schyns. 2001. “Bubbles: A Technique to Reveal the Use of Information in Recognition Tasks.” *Vision Research* 41 (17): 2261–71.

Carlson, T. A., Grootswagers, T., & Robinson, A. K. (2019). An introduction to time-resolved decoding analysis for M/EEG. *arXiv preprint arXiv:1905.04820*.

Dobs, Katharina, Leyla Isik, Dimitrios Pantazis, and Nancy Kanwisher. 2019. “How Face Perception Unfolds over Time.” *Nature Communications* 10 (1): 1258.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*, *18*(4), 203-210.

Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. *Scientific reports*, *9*(1), 6033.

Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., Gosselin, F., Hadley, D. M., Cavanagh, J. & Schyns, P. G., (2008). Classification images reveal the information sensitivity of brain voxels in fMRI.

Willenbockel, V., Gosselin, F. & Võ (submitted). Spatial frequency tuning for indoor scene categorization

Zhan, Jiayu, Robin A. A. Ince, Nicola van Rijsbergen, and Philippe G. Schyns. 2019. “Dynamic Construction of Reduced Representations in the Brain for Perceptual Decision Behavior.” *Current Biology: CB* 29 (2): 319–26.e4.

Annexe II : Article 7

Time course of the use of chromatic and achromatic facial information for sex categorization

Dupuis-Roy[1] N., Faghel-Soubeyrand[1], S., Gosselin[1], F.

1. Département de psychologie, Université de Montréal

Publié : Dupuis-Roy, N., **Faghel-Soubeyrand, S.**, & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43. https://doi.org/10.1016/j.visres.2018.08.004

Abstract

The most useful facial features for sex categorization are the eyes, the eyebrows, and the mouth. Dupuis-Roy et al. (2009) reported a large positive correlation between the use of the mouth region and rapid correct answers [Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of vision*, *9*(2), 10-10.]. Given the chromatic information in this region, they hypothesized that the extraction of chromatic and achromatic cues may have different time courses. Here, we tested this hypothesis directly: 110 participants categorized the sex of 300 face images whose chromatic and achromatic content was partially revealed through time (200 ms) and space using randomly located spatio-temporal Gaussian apertures (i.e. the Bubbles technique). This also allowed us to directly compare, for the first time, the relative importance of chromatic and achromatic facial cues for sex categorization. Results showed that face-sex categorization relies mostly on achromatic (luminance) information concentrated in the eye and eyebrow regions, especially the left eye and eyebrow. Additional analyses indicated that chromatic information located in the mouth/philtrum region was used earlier—peaking as early as 35 ms after stimulus onset—than achromatic information in the eye regions—peaking between 165-176 ms after stimulus onset, as was speculated by Dupuis-Roy et al. (2009). A non-linear analysis failed to support Yip and Sinha's (2002) proposal that processing of chromatic variations can improve subsequent processing of achromatic spatial cues, possibly via surface segmentation [Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*(8), 995-1003]. Instead, we argue that the brain prioritizes chromatic information to compensate for the sluggishness of chromatic processing in early visual areas, and allow chromatic and achromatic information to reach higher-level visual areas simultaneously.

Introduction

What are the respective roles of chromatic and achromatic cues in face-sex categorization? More specifically, which chromatic and achromatic facial cues does the brain use to perform this perceptual task, and what is the time course of their extraction? Research on face recognition has long underestimated the role of color, probably in part as a result of the popularity of earlier *edge-based* object recognition theories emphasizing the role of shape and spatial cues (Biederman and Ju, 1988; Marr, 1981). This view certainly seems difficult to reconcile with the emergence of trichromatic perception in primates (see Rowe, 2002). It is usually assumed that trichromatic perception had emerged through natural selection because it helped with the detection of certain fruits against approximately equiluminant green foliage (i.e. the foraging hypothesis, see Cain, Surridge & Mundy, 2003; Dominy & Luca, 2001). Evolutionary theorists have long speculated (see Allen, 1892) that once the appropriate biological apparatus for trichromatic perception had emerged through natural selection, it became useful in primate species for discriminating the spectral modulations on the facial skin of conspecifics, presumably for the purpose of discriminating emotional states, socio-sexual signals and threat displays (e.g. redder skin is related to higher testosterone and dominance level in male mandrills; ovulation in female chimpanzees is linked to redder skin; see Setchell & Dixson, 2001; Surridge, Osorio & Mundy, 2003). It was even recently argued that color vision in primates was originally selected specifically for social facial communication purposes (Changizi, Zhang & Shimojo, 2006; Hiramatsu, Melin, Allen, Dubuc & Higham, 2017).

Surprisingly, thus, it is only a few years ago that researchers began accumulating evidences that color is an important cue for face recognition in humans. Indeed, the chromatic properties of facial skin and features, particularly on the red-green axis, were shown to play an important role in facial categorization of social relevant information such as emotion (Young, Elliot, Feltman & Ambady, 2013; Benitez-Quiroz, Srinivasan & Martinez, 2018), attractiveness (Stephen, Oldham, Perrett & Barton, 2012; Fink, Grammer & Matts, 2006; Jones, Russel & Ward, 2014), health

(Stephen et al., 2009; Fink, Grammer & Matts, 2006; Stephen, Coetzee, Law Smith & Perrett, 2009; Jones, Porcheron, Sweda, Morizot & Russel, 2016; Russell et al., 2014; Russell et al., 2016), dominance (Elliott et al., 2010; Stephen et al., 2012; Mileva, Jones, Russell, Little, 2016), aggressiveness (Stephen et al., 2012), identity (Chang, Bao & Tsao, 2017; Nestor, Plaut, Behrmann, 2013), sex categorization (Nestor et al., 2008a; Nestor et al., 2008b; Dupuis-Roy et al., 2009; Dupuis-Roy et al., 2014; Russell, Kramer & Jones; 2017; Russell, 2009) and sex typicality (Fink, Grammer, Matts, 2006; Stephen and McKeegan, 2010). For instance, Stephen and McKeegan (2010) instructed participants to adjust the color of the lips along each axis of the CIELab colorspace to enhance the apparent attractiveness and sex typicality of faces. They found that women were perceived more feminine and attractive when red contrast was increased between the lips and the surrounding skin. Similar studies indicated that men are perceived as more masculine and dominant when they possess a darker and redder complexion across the entire face (Stephen et al., 2012; Jones, Russell & Ward, 2015; Stephen and McKeegan, 2010).

Nestor et al. (2008a) asked participants to categorize the sex of androgynous faces embedded in white chromatic noise, and used reverse correlation to reveal which color variations drove their perception of sex. They found that human observers mostly relied on color variations in the eyes, nose and mouth regions along the red-green axis of the CIELab color space. These results indicate that observers' memory representations for sex categorization contain chromatic elements; they do not indicate, however, that these color elements reflect real-world face-sex variations or, relatedly, that they contribute at all to correct face-sex categorization.

In fact, there is evidence that chromatic surface properties are important cues for face sex categorization (Dupuis et al., 2009; Hill, Bruce & Akamatsu., 1995; Kemp, Pile, White, Musselman, 1996). Nestor et al., (2008b) found that an algorithm designed to categorize the sex of frontal view color faces, reached a higher accuracy when all three color CIELab channels were available than when only luminance was available, suggesting that color conveyed sex-related information beyond that carried by brightness alone. A similar conclusion was reached independently by Dupuis-Roy et al. (2009; see Figure 3c below).

A hypothesis formulated by Yip and Sinha (2002, also see Nestor et al., 2008a) who have observed similar benefits of color in a face identification task, states that color could improves the segmentation of the facial features' surfaces and thus indirectly help the visual system to process the relevant facial cues. Dupuis-Roy et al. (2014) found further support for this hypothesis in a face-sex categorization task. They compared performances of human observers to that of an ideal observer designed to make use of all available information. Subjects had to categorize the sex of three sets of color and grayscale male and female faces: (1) a set of androgynous faces that varied only in terms of their real-life interattribute distances (i.e. all pairwise distances between the eyes, the nose and the mouth), (2) a set of faces that varied in all respects except that they all displayed the same androgynous set of interattribute distances and (3) a set of original faces. Overall, subjects obtained higher performances with the color faces than with their grayscale version, suggesting that they made correct use of face-sex color information in faces. Moreover, the authors observed higher efficiencies (i.e. ratio of human to ideal performance) in the color than in the grayscale conditions. This was the case even when faces only varied in terms of their real-world interattribute that is, when color was not task-relevant. Then, in this condition at least, the increase of efficiency must have originated from an improvement at a relatively low-level of visual processing. But if Yip and Sinha's low-level segmentation hypothesis is correct, color should impact facial processing early in the recognition process.

Gegenfurtner and Rieger (2000) found evidence for such a low-level mechanism for scene recognition. They employed a delayed match-to-sample task with three scene image conditions: 1) encoded and retrieved in luminance, 2) encoded in color and luminance and retrieved in luminance, and 3) encoded and retrieved in color and luminance. Scenes presented from 16 to 54 ms were better recognized when encoded in color and luminance compared to luminance alone, even if they were later retrieved in luminance. This is consistent with an early, low-level, segmentation process. Interestingly, scenes that were encoded *and* retrieved in color and luminance led to higher accuracy than scenes encoded in color and luminance and retrieved in

luminance only for presentation times of 64 ms. This suggests a high-level color (i.e. representational) advantage in addition to the low-level one.

In the face recognition literature, data supporting a fast, low-level advantage of color information remains scarce. Dupuis-Roy et al. (2009) gathered *indirect* evidence of an early extraction of color information in the context of face-sex categorization task. In this study, male and female color face images were spatially sampled with randomly generated *bubbles* masks—i.e. uniform gray masks punctured by randomly located small 2D Gaussian apertures. Multiple linear regression was used to assess the relationship between the position of these *bubbles* on the face, and the accuracy and latency of the responses. This first revealed that the eyes/eyebrows regions were linked to accurate sex categorization (see also Brown & Perrett, 1993; Russell, 2003; Yamaguchi, Hirukawa & Kanazawa, 1995). Of particular relevance here was the finding that the mouth region was significantly linked to *fast* accurate responses (but not to fast incorrect responses). In other words, subjects responded correctly and relatively fast on trials in which the mouth area of face stimuli tended to be revealed by bubbles. A statistical analysis of the face images revealed that the most discriminative information for the task within the mouth area was concentrated in the red-green color channel. Together, these results suggest that the processing of chromatic and achromatic facial cues follows a distinct time course, with color-based cues being processed earlier than luminance-based cues.

Here, we *directly* tested this hypothesis by examining the time course of the extraction of chromatic and achromatic cues in a face-sex categorization task. To do so, 110 participants were asked to categorize the sex of 900 unique stimuli created from a dataset of 300 real-life face images. The stimuli were made by sampling randomly the chromatic and the achromatic content of the face images through time and space with spatiotemporal Gaussian apertures. Computing the weighted sums of these spatiotemporal samples with accuracy on a trial-by-trial basis allows to determine the achromatic and chromatic voxels that were the most useful for categorizing the sex of a face through time. This experiment also gave us the opportunity to compare directly, for

the first time, the relative importance of chromatic and achromatic facial cues for sex categorization.

Methods

Participants

One hundred and ten healthy participants (54 men) with normal color vision and normal or corrected-to-normal visual acuity were recruited on the campus of the Université de Montréal. Participants were aged between 20 and 30 years. An informed consent was obtained prior to the experiment and a monetary compensation of 10$/h was provided upon its completion. All procedures were carried out with the approval of the Université de Montréal ethics committee.

Face database

We used the face database of Dupuis-Roy et al. (2009). It contains 300 Caucasian frontal-view face images (150 males and 150 females) with a neutral expression. Rotations, scalings, and translations in the image plane were applied to the face images in order to minimize the distance between handpicked landmarks around the eyes (four landmarks each), the eyebrows (2 landmarks each), the nose (four landmarks) and the mouth (four landmarks). Note that these so-called Procrustes transformations do not modify the relative distances between features.

Stimuli

The chromatic and achromatic content of a face image were revealed through Gaussian apertures randomly located in space (128x128 pixels for a given frame) and at random moments in the time (17-frames) dimension. This required four steps. A face image was first converted from its original RGB map to the HSV color map. Second, two volumes of 128 x 128 x 17 voxels, henceforth referred to as chromatic (Saturation, M1) and achromatic (Value, M2) volumes, with the same number of randomly located 3D Gaussian apertures were generated.

Each of these 3D Gaussian apertures (σ_{xy} = 0.15° of visual angle, σ_{time} = 23.53 ms) were scaled and truncated so that their probability density ranged from zero (at the tails ± 2.18 σ) to one (at the center). The spatial and temporal extents of the bubbles were selected to provide the resolution necessary to reveal the expected signals in the classification volumes. A spatial standard deviation of 0.15 deg of visual angle corresponds to a Full Width at Half Maximum (FWHM) of 0.35 deg of visual angle that is, approximately the period of the finest spatial information used by human observers to categorize the sex of faces (e.g. Willenbockel et al., 2010). A temporal standard deviation of 23.53 ms corresponds to a FWHM of 54.75 ms that is, slightly less than the smallest temporal period that has been observed in similar experiments (e.g. Vinette, Gosselin & Schyns, 2004; Blais et al., 2013). Third, the following calculations were performed on the 17 frames of the H (hue), S (saturation) and V (value) channels of a given face image:

Eq 1 :

$$H'_i = H$$

$$S'_i = M1_i .* S$$

$$V'_i = M2_i .* (V-\mu)+\mu \ ,$$

where i=[1:17]; H, S and V are the original channels of a given face image; H', S' and V' are the resulting sampled volumes (128 x 128 x 17 voxels); and μ is the average of all pixels in V. In conjunction with the H channel, each bubble of the chromatic volume M1 reveals hue information with the maximum saturation at its centre. Each bubble of the achromatic bubbles volume M2 reveals the maximum value at its centre. When bubbles of the chromatic and achromatic volume do not overlap, the areas revealed through the chromatic bubbles have approximately the same average luminance (± 5 cd/m^2, which represents 3.7% of the full range), whereas the areas revealed through the achromatic bubbles have a null saturation and thus no hue. When bubbles of the chromatic and achromatic volumes overlap perfectly, they reveal all the information to its full contrast. Consequently, the degree of spatial overlapping between both M1 and M2 bubbles volume establish how much of the whole HSV face information is revealed. Finally, the H', S' and V' volumes are converted to RGB values. Figure 1a illustrates one

iteration of the sampling process that was repeated for each time frame, 17 times. Note that face images were set at a low resolution of 128 x 128 pixels to keep statistical power high while exploring two tridimensional search spaces.

The experimental programs were run on four Macintosh computers in the Matlab (Mathwork Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The computers' high-resolution CRT monitors were set to display 1024 by 768 pixels at a refresh rate of 85 Hz. The relationship between RGB values and luminance levels (measured with a Samsung SyncMaster 753df photometer) was computed for each color channel independently; the three best-fitted gamma functions were then used to convert each image of the face database into HSV space. Participants were seated in a dim-lighted room. Their viewing distance was maintained by a chin-rest so that the stimuli interpupillary distance was 1.03 deg of visual angle (face width was 3.28 deg of visual angle) as in Dupuis-Roy et al. (2009), which is equivalent to viewing a real face from approximately 3.50 m. Thus proximal stimuli fell within the fovea—a ~5 deg area on the retina where fine details and colors are best distinguished. Although the relatively small angular size of our face images might have promoted the use of information at slightly lower spatial frequencies than in the typical face recognition experiment, it probably had a limited impact on our spatio-temporal results (Loftus & Harley, 2005; Näsänen, 1999; Willenbockel et al., 2010; but see Yang, Shafai & Oruq, 2015).

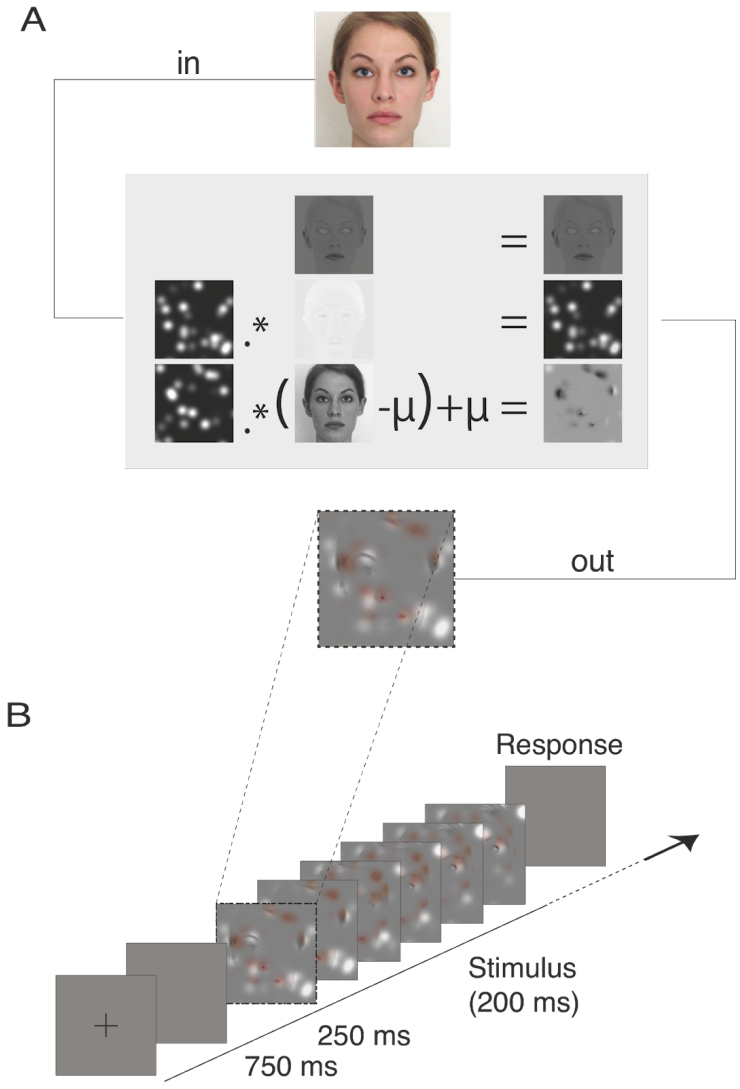


Figure 9.1. A) Illustration of the construction of one of the 17 frames of a face stimulus. First, a face image is converted from its original RGB color map to the HSV color map. Second, two volumes of Gaussian apertures randomly located in space (128 x 128 pixels) and at random moments in time (17 frames) were generated (only one frame shown per volume). One volume is used to samples the S channel (saturation) of the face image, while the other volume is used to sample the V channel (value). The H (hue) channel of the image is fully revealed in the face stimulus. Each volume had the same number of randomly located 3D Gaussian apertures. Finally, the H, S and V volumes are converted to RGB volumes. B) On each trial, a fixation cross is presented at the center of the computer monitor for 750 ms; immediately after a homogeneous gray field is presented for 250 ms; immediately after the 17 frames of a face stimulus are shown at a rate of 85 Hz (for a total duration of 200 ms); immediately after a homogenous gray field is presented until a response is provided by the participant.

Procedure

First, participants were screened for possible color vision deficiencies with the Ishihara Color Test (Ishihara, 1936). Then, they were brought in a dark lighted experimental room where they completed three blocks of 300 trials in each of which a different face from the face database was shown, in a random order, partially revealed in space (128x128 pixels for a given frame) and time (200 ms) through 3D bubble masks. Face photos were rotated 180 deg about its vertical axis with a 0.5 probability on every trial to eliminate possible information asymmetries (e.g. illumination differences). The sequence of visual events in a given trial unfolded as follows (see Figure 1b): A black fixation cross was shown at the center of the computer monitor against a uniform gray background for 750 ms; a uniform grey field immediately followed for 250 ms; a video containing the 17 frames of a stimulus immediately followed at the center of the screen against a uniform gray background for 200 ms (i.e., 17 frames presented at 85 Hz); and, finally, a uniform grey field immediately followed and remained until the participant had indicated the sex of the sampled face by pressing a labeled computer keyboard key. No feedback was provided. The number of bubbles was adjusted on a trial-by-trial basis with Quest (Watson & Pelli, 1983) to maintain 75% of correct responses.

Results

Participants required an average of 97.23 Gaussian apertures (SD = 35.95) to reach the target of 75% of correct responses. Four participants were excluded from the rest of the data analyses because the number of bubbles they required was at least three standard deviations above the group average. Thus the following analyses were carried out on a grand total of 106 subjects (for a total of 95,400 trials).

Linear classification volumes

Two weighted sums were conducted for every participant to determine the achromatic and chromatic voxels that were the most useful for categorizing the sex of a face through time. We summed the volumes of chromatic Gaussian apertures weighted by the participant's response accuracy transformed in z-scores on a trial-by-trial basis; and we summed the volumes of achromatic Gaussian apertures and the participant's response accuracy transformed in z-scores on a trial-by-trial basis. The resulting classification volumes (128 x 128 x 17) revealed how each participant extracted achromatic and chromatic face-sex information through time.

To assess the relative use of achromatic and chromatic cues, we summed, for each participant, the unsmoothed standardized voxels found within the left eye, the right eye and the mouth areas—defined anatomically—for both the achromatic and chromatic individual classification volumes. This led to two sums of z-scores per participant that is, one for the achromatic and another for the chromatic classification volume. The values obtained for the two classification volumes were then compared with a two-tailed paired sample t-test. This revealed that achromatic facial information was 4.01 times more potent than that of chromatic facial information for sex-categorization ($M_{achromatic}$=0.75, SD=0.95; $M_{chromatic}$=0.19, SD=0.98; t(105)=4.32, p<.001, Cohen's d=.58).

The group-average chromatic and achromatic classification volumes were also computed by summing the appropriate individual classification volumes. These group classification volumes were then smoothed with the same 3D Gaussian kernel as the one used in the experiment (σ_{xy}=0.15° of visual angle, σ_{time}=23.53 ms) and finally transformed into z-scores. This transformation was based on a bootstrap estimation (278 528 voxels) of the null hypothesis population mean and standard deviation (e.g., Efron & Tibrishiani, 1986; Chauvin et al., 2005**)**.

The remaining statistical analyses performed on the group classification volumes were restricted to the face area, i.e. a region of 3,469 pixels per frame that excluded the hair and other external facial cues. Any significant positive local divergence from uniformity in our group classification volumes would indicate that the corresponding part of the stimuli led to more accurate responses. A one-tail Pixel test was done on each z-scored group classification volume

(S_r= 58,973 voxels , Z_{thresh} = 3.05, $p<0.05$). This test — which was derived from the Random Field Theory — predicts what statistical values should be expected (whether Z, F or t-stat) when a tensor of a given random distribution is convolved with a given kernel. It thus corrects for multiple comparisons while taking the spatial correlation inherent to smoothed classification volumes into account. Further details about this test are provided in Chauvin et al. (2005).

[Insert Figure 2 about here]

The use of spatial features

Each blob of Figure 3b represents a spatiotemporal cluster of significant voxels that was collapsed on the temporal dimension. The blobs are depicted over the average face to help with interpretation. The results show that the eyes and eyebrows, especially the left eye and eyebrow, in the achromatic classification volume (green shaded areas) are the most important facial regions for sex categorization followed by the mouth region (i.e. the philtrum and the mentolabial sulcus) in the chromatic classification volume (red shaded areas).

We tested the apparent asymmetry between use of the left and right eye with a two-way repeated measure ANOVA (2 eye regions x 17 frames) over the two classification volumes. We found a significant main effect for the eye regions in the achromatic classification volume—a greater use of the left eye/eyebrow than the right from the observer's point of view ($F(1,105)=4.36$, $p=0.04$, $M_{left}=0.20$, $SD_{left}=0.96$, $M_{right}=0.13$, $SD_{right}=1.01$). Importantly, this asymmetry in the use of facial information cannot be attributed to physical differences between the left and right side of face stimuli because the partially revealed face images were mirror-reversed about their vertical axis with a probability of 0.5 on each trial. There was no such effect in the chromatic classification volume ($F(1,105)=0.64$, $p=0.43$).

Figure 3c, adapted from Dupuis-Roy et al., (2009, Figure 3), shows the available chromatic and achromatic information in the face database that we used. Each face images was first converted to *Lab* color space: *L* corresponds to the light-dark—or achromatic—color map, *a* to the red-green color map, and *b* to the yellow-blue color map. Then, we computed an index of the information available at each pixel of each color map to discriminate the sex of a face. More

specifically, we calculated *d'*s—the distance between the mean of the distributions of this pixel's values for male and for female faces in standard deviation units. The three *d'* maps are represented as contour plots in Figure 3c. Warm colors indicate regions where men are lighter, redder or yellower than women and, cold colors, regions where men are darker, greener or bluer than women. To help with interpretation, the contour plots are superimposed to the face average. Thick white dotted lines delineate the main significant achromatic (see light-dark color map) and chromatic (see red-green and yellow-blue color maps) blobs in our classification volumes.

The use of spatial features through time

At the temporal level, the achromatic clusters in both eyes are significant for the entire sampling period, i.e. from 12 to 200 ms after stimulus onset, whereas the chromatic clusters in the mouth region are significant from 12 to 24 ms and from 165 to 200 ms after stimulus onset (see Figure 2). The time at which clusters are significant, however, only provides an incomplete measure of the time course of information extraction because it lacks granularity and also because of the large differences in z-scores between the chromatic and achromatic classification volumes. To overcome these shortcomings, we divided the maximum z-score found in the eye and mouth areas defined anatomically at each point in time by the overall maximum z-score found in the appropriate classification volume. A bootstrap analysis was also conducted to simulate the distribution of maximum z-scores found in each region of interest given the null hypothesis (on a total of 278,528 voxels), and to compute a statistical threshold.

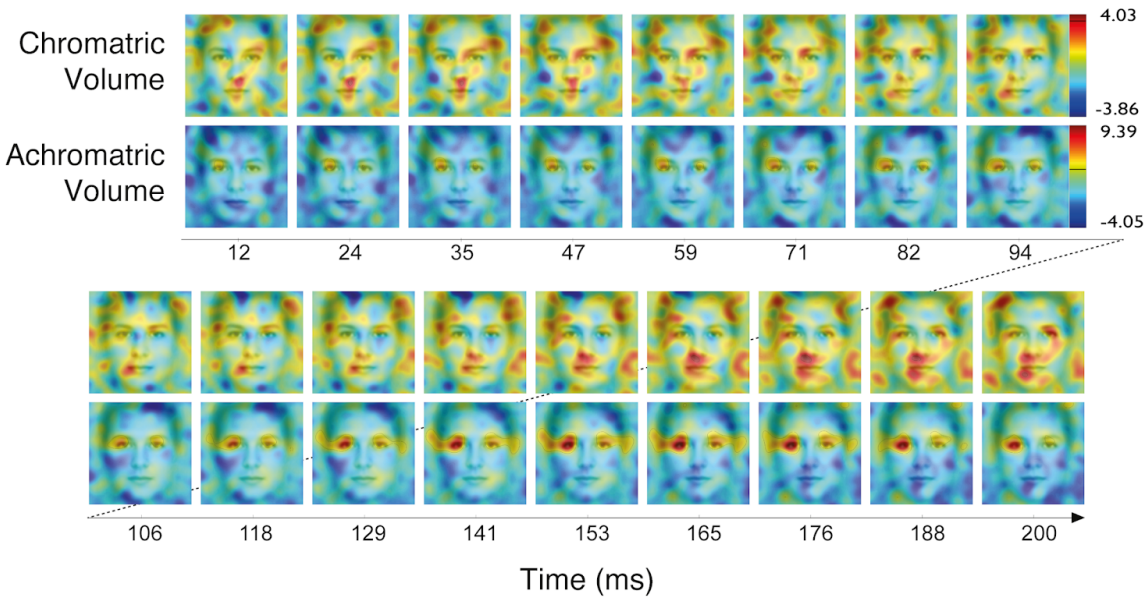


Figure 9.2. Linear classification volumes for facial sex categorization. A weighted sum of the volumes of chromatic Gaussian apertures (explanatory variables) by the participant's response accuracy (dependant variable) was computed. The same procedure was also done between the volumes of achromatic Gaussian apertures and the participant's response accuracy. Spatio-temporal regions containing significant regression coefficients for chromatic (2nd and 4th row) and chromatic face information (1st and 3rd row) are delimited by thin black lines (pixel-test, $p<.05$). The frames of the classification volumes were superimposed on the average face to help with interpretation.

The resulting curves provide the time course of chromatic information extraction in the mouth region (red curve) as well as the time course of achromatic information extraction in the left (green curve) and in the right eye (dotted green curve) regions from the point of view of the observer (Figure 3a), factoring out the absolute usefulness of achromatic and chromatic information. Note that the maximum of the curves within each classification volume was set to 1. The colored circles indicate time points that attain statistical significance ($p<0.05$). Chromatic information extraction within the mouth region peaks at 35 ms and then again at 176 ms after the

stimulus onset, whereas the achromatic information extraction within the left and right eye regions respectively peaks at 165 and 176 ms after the stimulus onset.

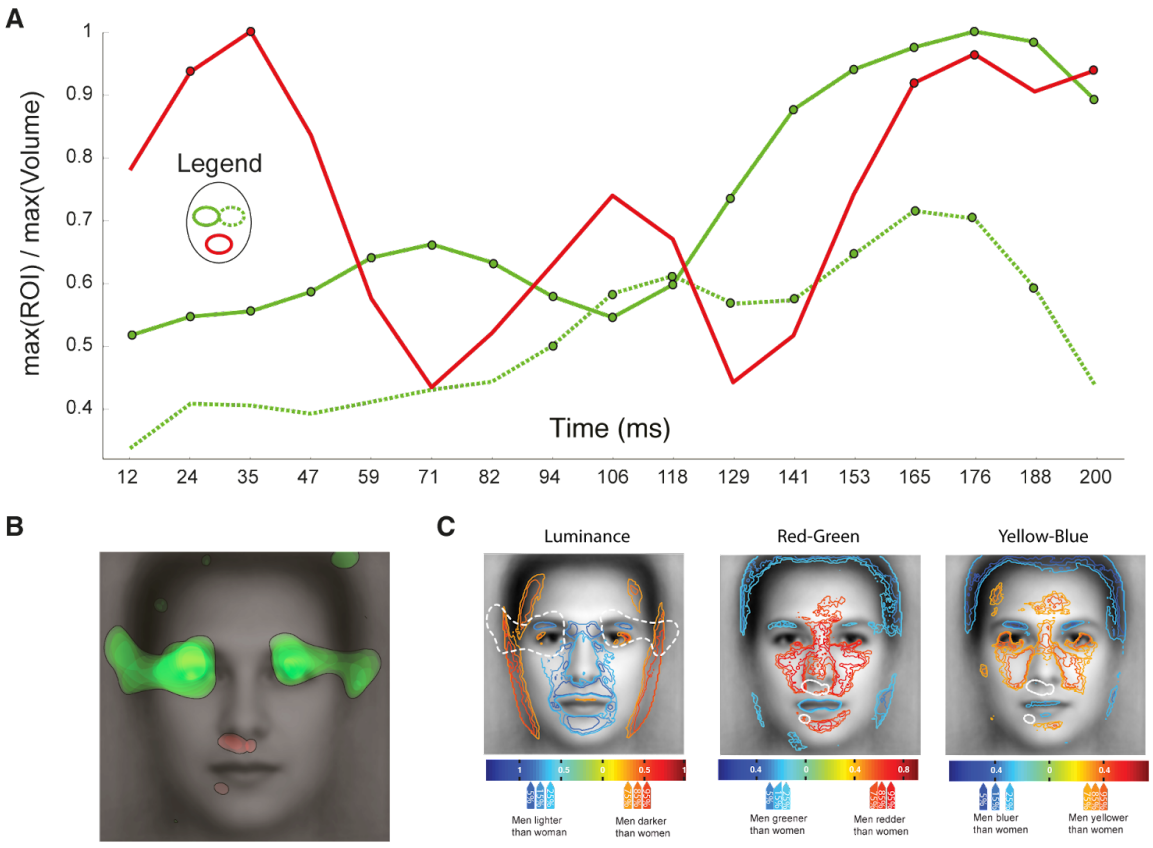


Figure 9.3. Spatiotemporal use of chromatic and achromatic facial cues for facial sex categorization. A) The solid red curve represents the time course of the use of chromatic information in the mouth region; the green curves represent the time course of the use of the left (solid line) and right eye region (dotted line) from the observer's perspective. Open circles indicate statistically significant curve segments ($p<.05$). B) Green and red blobs represent, respectively, the spatio-temporal regions that contained regression coefficients significantly greater than zero in the achromatic and chromatic classification volumes ($p<.05$). These blobs were collapsed on the temporal dimension and superimposed on the average face to help with interpretation. C) Contour plots of the available color/luminance information for face-sex discrimination are superimposed to the average face. The contour plot summarizes the spatial modulation of available information (d's) in the dark-light (1st image from the left), red-green (2nd) and yellow-blue (3rd) channels. Warm colors were used for regions where men are lighter, redder or yellower than women; and cold colors for regions where men are darker, greener or

bluer than women. Clusters significantly correlated with accurate achromatic (see white dotted line in the light-dark color map) and accurate chromatic responses (full white lines in the red-green and yellow-blue color maps) from Figure 3B reveal which specific available information may be used by observers.

Yip and Sinha's segmentation hypothesis

Yip and Sinha (2002) proposed that processing of chromatic variations, whether task-relevant or not, could facilitate the segmentation and, therefore, subsequent processing of achromatic spatial features. Contrary to this hypothesis, we did not find any facial region in which significant voxels in the chromatic classification volume peaked *before* significant voxels in the achromatic classification volume. Our linear analysis, however, can only reveal cases of chromatic segmentation and subsequent use of achromatic cues occuring systematically at the same latencies after stimulus onset. To test more thoroughly the chromatic-before-achromatic hypothesis, we first selected, for each participant, the trials that revealed chromatic voxels at least 5 frames (corresponding to the FWHM of the gaussian aperture on the temporal dimension) before achromatic voxels in each of the three facial regions of interest (M=272.17 trials per participant), and then computed the mean accuracy. One-tailed paired sample t-tests between these mean accuracies and the ones obtained similarly for the alternative achromatic-before-chromatic hypothesis failed to reveal any significant difference (left eye: $t(105)=-0.33$, $p=0.74$; $M_{chromatic\text{-}before\text{-}achromatic}=0.756$, $M_{achromatic\text{-}before\text{-}chromatic}=0.757$; right eye: $t(105)=0.04$, $p=0.97$; $M_{chromatic\text{-}before\text{-}achromatic}=0.755$, $M_{achromatic\text{-}before\text{-}chromatic}=0.755$; mouth: $t(105)=1.83$, $p=0.07$; $M_{chromatic\text{-}before\text{-}achromatic}= 0.758$, $M_{achromatic\text{-}before\text{-}chromatic}= 0.752$).

Discussion

Which chromatic and achromatic facial cues does the brain use to perform sex categorization, and what is the time course of their extraction? Research on face recognition has long underestimated the role of color. Recent studies showed that the chromatic properties of

facial skin and features, particularly on the red-green axis, play an important role in many socially relevant facial judgments, including sex categorization (e.g. Nestor et al., 2008a; Nestor et al., 2008b; Dupuis-Roy et al., 2009; Dupuis-Roy et al., 2014; Russell, Kramer & Jones; 2017; Russell, 2009). Studies are also starting to unravel how chromatic and achromatic information interact through time during face categorization. Findings by Dupuis-Roy et al. (2009), for example, suggest that the extraction of chromatic cues from the mouth region precedes the extraction of achromatic cues from the eye regions (see also Gegenfurtner and Rieger, 2000; Oliva & Schyns, 2000).

Here, we examined directly the time course of local chromatic and achromatic face information extraction during a face-sex categorization task in more than 100 observers. We found that achromatic information was about four times more useful for sex categorization than chromatic information. We also showed that the most useful chromatic facial information was concentrated in the mouth region (i.e. the philtrum and the mentolabial sulcus) of the face, and that the most useful achromatic information was concentrated in the eye and eyebrow regions. Furthermore, we revealed, for the first time, that chromatic information extraction peaks as early as 35 ms while achromatic information extraction peaks later, between 165 and 176 ms. Finally, we tested, and refutated, one possible explanation for this chromatic-before-achromatic-information extraction, namely that it could reflect a coarse chromatic segmentation of facial features prior to their fine achromatic analysis (Yip & Sinha, 2002).

The spatial distribution of chromatic and achromatic face-sex cues

The magnitude of z-scores obtained in the group achromatic classification volume were approximately four times greater than that of the z-scores obtained in the group chromatic classification volume. In other words, achromatic information is much more useful to human observers to categorize the sex of faces than chromatic information. This is in line with previous studies on face image statistics which revealed that there is little information *available* to

categorize sex in the chromatic compared to the achromatic channel (Figure 3c; see also Kemp et al.,1996; Nestor & Tarr, 2008b, Russell, 2003; Russell, 2009; Russell, 2015; Dupuis-Roy et al., 2009; Dupuis-Roy et al., 2014). This is also compatible with previous results indicating that human observers do represent chromatic information for sex categorization (Nestor & Tarr, 2008a), and that they use this scarce information quite efficiently (Dupuis-Roy et al., 2014).

We replicated the findings that the eye, eyebrow and mouth regions are the most prominent face-sex cues for human observers (e.g. Dupuis-Roy et al., 2009; Brown & Perrett, 1993; Nestor & Tarr, 2008a; Roberts, 1988; Yamaguchi, Hirukawa, & Kanazawa, 1995; Schyns, Bonnar, & Gosselin, 2002; Gosselin & Schyns, 2001; Burton, Bruce, & Dench, 1993; Russell, 2003). We further discovered that it is the chromatic information within the philtrum and mentolabial sulcus areas, and the achromatic information within the eye and eyebrow areas specifically which are the most useful to human observers. In the chromatic classification volume, the significant spatiotemporal clusters in the philtrum and mentolabial sulcus areas, as well as the marginally significant cluster in the left oral commissure area from the observer's perspective (see Figure 2, frames 7-12; maximum z-score=2.936, p=0.10) all match real-world chromatic sexual dimorphisms. As shown in Figure 3c, these sexual dimorphisms are concentrated in the red-green face color maps: e.g., men tend to have greener lips and redder skin in the philtrum and mentolabial sulcus areas. In the achromatic classification volume, the significant clusters in the eye and eyebrow regions contain all featural achromatic sexual dimorphisms. Dupuis-Roy et al. (2009) had speculated that this was the case based on the use of spatial information—achromatic and chromatic information combined—by human observers and the available information in these different facial regions but this had never been shown empirically.

Several prior studies reported an asymmetric use of information favoring the left over the right eye and eyebrow region from the observer's perspective, both using the Bubbles procedure (e.g. Faghel-Soubeyrand, Dupuis-Roy & Gosselin, submitted; Gosselin & Schyns, 2001; Ince et al., 2016; Rousselet et al., 2014; Schyns, Bonnar & Gosselin, 2002; Vinette, Gosselin & Schyns,

2004) and other experimental procedures, such as divided visual field, chimeric faces or gaze-tracking (e.g. Bourne, 2008; Butler et al., 2005; Jansari, Rodway, & Goncalves, 2011; Innes, Burt, Birch & Hausmann, 2015; Yovel, Timbini & Branman, 2008). Here, we extend these findings by showing for the first time that this asymmetry is driven exclusively by achromatic cues, and that it is present as soon as 12 ms after stimulus onset and lasts at least 200 ms.

Time course of chromatic and achromatic cues

Our experiment showed that achromatic and chromatic facial information are extracted according to distinct time courses. We examine these time courses in the next paragraphs.

The use of achromatic information within the left and right eye regions, respectively, attain local maxima at ~71 ms and at ~118 ms, and then culminate at ~176 ms and ~165 ms. It is tempting—but it would be an error—to associate the peaks of these time courses with the N170, an event related-potential component associated to face processing that occurs about 170 ms after stimulus onset and that is widely believed to originate from the inferotemporal cortex (IT, Bentin, 1996; for a review see Rossion, 2014). The moment at which information is processed is function of both the time window of extraction and the processing speed of the visual system. Indeed, we know that it takes at least 45 ms for a neural signal to travel from the photoreceptors to V1, and at least an additional 60 ms to reach IT (Bullier, 2001). This implies that information impinging the retina during the 165-176 time window after the first frame of our stimuli is actually processed more than 105 ms later in IT, that is more than 270-281 ms after stimulus onset. These concepts—processing and extraction—are often confused in the literature (see also VanRullen, 2011); this can lead to processing-oriented interpretations, when in fact the results are about extraction, and vice-versa.

The local maxima of the achromatic time courses (~71 ms and ~118 ms, respectively, for the left- and right-eye regions) are, however, consistent with the timing of the N170. Smith et al. (2004) showed that this early face-sensitive brain areas are mainly driven by the contralateral eye, irrespective of the categorization task (see also Lissa et al., 2014; Rousselet et al., 2014; Ince et

al., 2016), whereas selective attention to diagnostic information for categorizing stimuli—achromatic eye information and chromatic mouth information for the task at hand—correlates with late brain activity in the 250-300ms time window (see also Hachemi et al., 2018). This late brain activity could reflect the processing of the peaks in the achromatic time courses (as well as the late burst in the chromatic time course).

The most important contribution of this study is without a doubt the uncovering of the time course of the use of isoluminant chromatic face information. Chromatic cues from the mouth area are extracted in two successive bursts, the first of which happens as early as 35 ms after stimulus onset while the second one happens around 176 ms. Importantly, the early chromatic burst occurs about 135 ms before the peak of the extraction of achromatic cues, and 35 ms earlier than its first local maxima. This offers direct support for the hypothesis formulated by Dupuis-Roy et al. (2009) that chromatic information in the mouth area is extracted prior to achromatic information in the eye areas. It seems, however, at odd with several studies that have reported that the parvocellular system—specialized in the processing of fine chromatic information—is generally slower than the magnocellular system—specialized in the processing of coarse achromatic information (Merigan and Maunsell, 1993). For example, Nowak et al. (1995; see also Maunsell et al., 1999) found that V2 colour cells were activated 10-20 ms after V2 non-colour cells when stimulated with simple spots of light. But this story appears to change with more complex face stimuli. Edwards et al. (2003) examined the effect of the presence of color in face stimuli on the response of IT neurons in monkeys. They found no evidence to support the hypothesis that chromatic information about faces is delayed with respect to achromatic information in IT. They proposed that the brain compensates for the relative sluggishness of color processing in early visual areas such as V1 and V2 to allow chromatic and achromatic information to reach higher-level visual areas simultaneously. Our results suggest that the brain could achieve this goal by prioritizing the extraction of chromatic facial information ~135 ms earlier than achromatic information.

A proposal by Yip and Sinha (2002) suggests that the brain could actually take advantage of the processing of chromatic before achromatic information in low- to mid-level ventral stream brain areas—prior to the integration of chromatic and achromatic information. They proposed that chromatic variations, whether task-relevant or not, contribute to the segmentation of surfaces in different regions and, thus, improve the subsequent processing of achromatic spatial cues (see also De Valois & Switkes, 1983; Nestor et al., 2008b; Dupuis-Roy et al., 2014). Our results, however, failed to support this hypothesis. First, our linear classification volumes contain no facial region for which the extraction of chromatic information precedes the extraction of achromatic information. Second, and more conclusively, trials on which chromatic information was presented before achromatic information within a given facial region of interest—irrespective of the interval between the sources of information—did not lead to greater performances than trials on which achromatic information was presented before chromatic information.

The integration of chromatic and achromatic facial cues could occur in a subset of four recently discovered color-selective brain areas anterior to the V4/hV4 color area (Zeki et al.,1991; Brewer et al., 2005) and "sandwiched" between face and place-sensitive regions along IT of humans and primates (Chang, Bao & Tsao, 2017; Lafer-Sousa, Conway & Kanwisher, 2013; Lafer-Sousa, Conway & Kanwisher, 2016). These hue-sensitive modules tend to have a higher overlap with face-selective patches as they go through a posterior to anterior axis of IT (Lafer-Sousa, Conway & Kanwisher, 2013). Interestingly, the most anterior of these color modules, called the "anterior medial color" region, is highly responsive to chromatic information about faces and bodies (Chang, Bao & Tsao, 2017; see also Simmons et al., 2007). However, more research will be required to examine when and where in the brain are processed the chromatic and achromatic face-sex information extracted in different time windows.

References

Allen, G. (1892). *The colour-sense: its origin and development: an essay in comparative psychology* (Vol. 10). London: K. Paul, Trench, Trübner.

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: new results and a review. *European Journal of Neuroscience*, *12*(1), 172-193.

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2018). Facial color is an efficient mechanism to visually transmit emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3581–3586. http://doi.org/10.1073/pnas.1716084115

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, *20*(1), 38-64.

Blais, C., Arguin, M. & Gosselin, F. (2013). Human visual processing oscillates: Evidence from a classification image technique. Cognition, 128, 353-362

Bourne, V. J. (2008). Chimeric faces, visual field bias, and reaction time bias: Have we been missing a trick?. *Laterality*, *13*(1), 92-103.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433-436.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature neuroscience*, *8*(8), 1102.

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns?. *Neuropsychologia*, *43*(1), 52-59.

Caine, N. G., Surridge, A. K., & Mundy, N. I. (2003). Dichromatic and trichromatic Callithrix geoffroyi differ in relative foraging ability for red-green color-camouflaged and non-camouflaged food. *International Journal of Primatology*, *24*(6), 1163-1175.

Chang, L., Bao, P., & Tsao, D. Y. (2017). The representation of colored objects in macaque color patches. *Nature communications*, *8*(1), 2064.

Changizi, M. A., Zhang, Q., & Shimojo, S. (2006). Bare skin, blood and the evolution of primate colour vision. *Biology letters*, *2*(2), 217-221.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of vision*, *5*(9), 1-1.

De Valois, K. K., & Switkes, E. (1983). Simultaneous masking interactions between chromatic and luminance gratings. *JOSA*, *73*(1), 11-18.

Dominy, N. J., & Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature*, *410*(6826), 363.

Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering sex discrimination cues in a realistic setting. *Journal of Vision*, *9*(2), 10-10.

Dupuis-Roy, N., Fiset, D., Dufresne, K., Caplette, L., & Gosselin, F. (2014). Real-world interattribute distances lead to inefficient face sex categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1289.

Edwards, R., Xiao, D., Keysers, C., Földiák, P., & Perrett, D. (2003). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *Journal of Neurophysiology*, *90*(2), 1245-1256.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, 54-75.

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General*, *139*(3), 399.

Faghel-Soubeyrand, S., Dupuis-Roy N., & Gosselin, F. (submitted). Inducing the use of right-eye face information enhances face-sex categorization performance.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, *27*(6), 433-442.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805-808.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*(6), 878-892.

Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the sex and race of faces: The role of shape and colour. *Proceedings of the Royal Society of London B: Biological Sciences*, *261*(1362), 367-373.

Hiramatsu, C., Melin, A. D., Allen, W. L., Dubuc, C., & Higham, J. P. (2017). Experimental evidence that primate trichromacy is well suited for detecting primate social colour signals. In *Proc. R. Soc. B* (Vol. 284, No. 1856, p. 20162458). The Royal Society.

Ince, R. A., Jaworska, K., Gross, J., Panzeri, S., Van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex*, *26*(11), 4123-4135.

Innes, B. R., Burt, D. M., Birch, Y. K., & Hausmann, M. (2016). A leftward bias however you look at it: Revisiting the emotional chimeric face task as a tool for measuring emotion lateralization. *Laterality: Asymmetries of Body, Brain and Cognition*, *21*(4-6), 643-661.

Ishihara, S. (1936). series of plates designed as tests for colour-blindness.

Jansari, A., Rodway, P., & Goncalves, S. (2011). Identifying facial emotions: Valence specific effects and an exploration of the effects of viewer sex. *Brain and Cognition*, *76*(3), 415-423.

Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics alter biologically-based factors of beauty: Evidence from facial contrast. *Evolutionary Psychology*, *13*(1), 147470491501300113.

Jones, A. L., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2016). Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image*, *17*, 57-66.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, *25*(1), 37-52.

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face-and place-selective regions in humans, as in macaques. *Journal of Neuroscience*, *36*(5), 1682-1697.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away?. *Psychonomic Bulletin & Review*, *12*(1), 43-65.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*(5233), 102.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, *207*(1167), 187-217.

Maunsell, J. H., Ghose, G. M., Assad, J. A., McADAMS, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual neuroscience*, *16*(1), 1-14.

Mileva, V. R., Jones, A. L., Russell, R., & Little, A. C. (2016). Sex differences in the perceived dominance and prestige of women with and without cosmetics. *Perception*, *45*(10), 1166-1183.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision research*, *39*(23), 3824-3833.

Nestor, A., & Tarr, M. J. (2008a). Sex recognition of human faces using color. *Psychological Science*, *19*(12), 1242-1246.

Nestor, A., & Tarr, M. J. (2008b). The segmental structure of faces and its use in sex recognition. *Journal of vision*, *8*(7), 7-7.

Nestor, A., Plaut, D. C., & Behrmann, M. (2013). Face-space architectures: Evidence for the use of independent color-based features. *Psychological science*, *24*(7), 1294-1300.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual neuroscience*, *12*(2), 371-384.

Nunn, J. A., Gregory, L. J., Brammer, M., Williams, S. C., Parslow, D. M., Morgan, M. J., ... & Gray, J. A. (2002). Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. *Nature neuroscience*, *5*(4), 371.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive psychology*, *41*(2), 176-210.

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

Rowe, M. H. (2002). Trichromatic color vision in primates. *Physiology*, *17*(3), 93-98.

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in cognitive sciences*, *18*(6), 310-318.

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*(9), 1093-1107.

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, *38*(8), 1211-1219.

Russell, R., Sweda, J. R., Porcheron, A., & Mauger, E. (2014). Sclera color changes with age and is a cue for perceiving age, health, and beauty. *Psychology and aging*, *29*(3), 626.

Russell, R., Porcheron, A., Sweda, J. R., Jones, A. L., Mauger, E., & Morizot, F. (2016). Facial contrast is a cue for perceiving health from the face. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1354.

Russell, R., Sweda, J. R., Porcheron, A., & Mauger, E. (2014). Sclera color changes with age and is a cue for perceiving age, health, and beauty. *Psychology and aging*, *29*(3), 626.

Russell, R., Kramer, S. S., & Jones, A. L. (2017). Facial Contrast Declines with Age but Remains Sexually Dimorphic Throughout Adulthood. *Adaptive Human Behavior and Physiology*, 1-11.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological science*, *13*(5), 402-409.

Setchell, J. M., & Dixson, A. F. (2001). Changes in the secondary sexual adornments of male mandrills (Mandrillus sphinx) are associated with gain and loss of alpha status. *Hormones and Behavior*, *39*(3), 177-184.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802-2810.

Stephen, I. D., Smith, M. J. L., Stirrat, M. R., & Perrett, D. I. (2009). Facial skin coloration affects perceived health of human faces. *International journal of primatology*, *30*(6), 845-857.

Stephen, I. D., Coetzee, V., Smith, M. L., & Perrett, D. I. (2009). Skin blood perfusion and oxygenation colour affect perceived human health. *PLos one*, *4*(4), e5083.

Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception*, *39*(8), 1104-1110.

Stephen, I. D., Oldham, F. H., Perrett, D. I., & Barton, R. A. (2012). Redness enhances perceived aggression, dominance and attractiveness in men's faces. *Evolutionary Psychology*, *10*(3), 147470491201000312.

Surridge, A. K., Osorio, D., & Mundy, N. I. (2003). Evolution and selection of trichromatic vision in primates. *Trends in Ecology & Evolution*, *18*(4), 198-205.

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

Vinette, C., Gosselin, F. & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes! Cognitive Science, 28, 289-301

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics*, *33*(2), 113-120.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., ... & Gosselin, F. (2010). Does face inversion change spatial frequency tuning?. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 122.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human perception and performance*, *19*(4), 899.

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*(8), 995-1003.

Yang, N., Shafai, F., & Oruc, I. (2014). Size determines whether specialized expert processes are engaged for recognition of faces. *Journal of vision*, *14*(8), 17-17.

Young, S. G., Elliot, A. J., Feltman, R., & Ambady, N. (2013). Red enhances the processing of facial expressions of anger. *Emotion*, *13*(3), 380.

Yovel, G., Tambini, A., & Brandman, T. (2008). The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. *Neuropsychologia*, *46*(13), 3061-3068.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of neuroscience*, *11*(3), 641-649.

Figure 1. A) Illustration of the construction of one of the 17 frames of a face stimulus. First, a face image is converted from its original RGB color map to the HSV color map. Second, two volumes of Gaussian apertures randomly located in space (128 x 128 pixels) and at random moments in time (17 frames) were generated (only one frame shown per volume). One volume is used to samples the S channel (saturation) of the face image, while the other volume is used to sample the V channel (value). The H (hue) channel of the image is fully revealed in the face stimulus. Each volume had the same number of randomly located 3D Gaussian apertures. Finally, the H, S and V volumes are converted to RGB volumes. B) On each trial, a fixation cross is presented at the center of the computer monitor for 750 ms; immediately after a homogeneous gray field is presented for 250 ms; immediately after the 17 frames of a face stimulus are shown at a rate of 85 Hz (for a total duration of 200 ms); immediately after a homogenous gray field is presented until a response is provided by the participant.

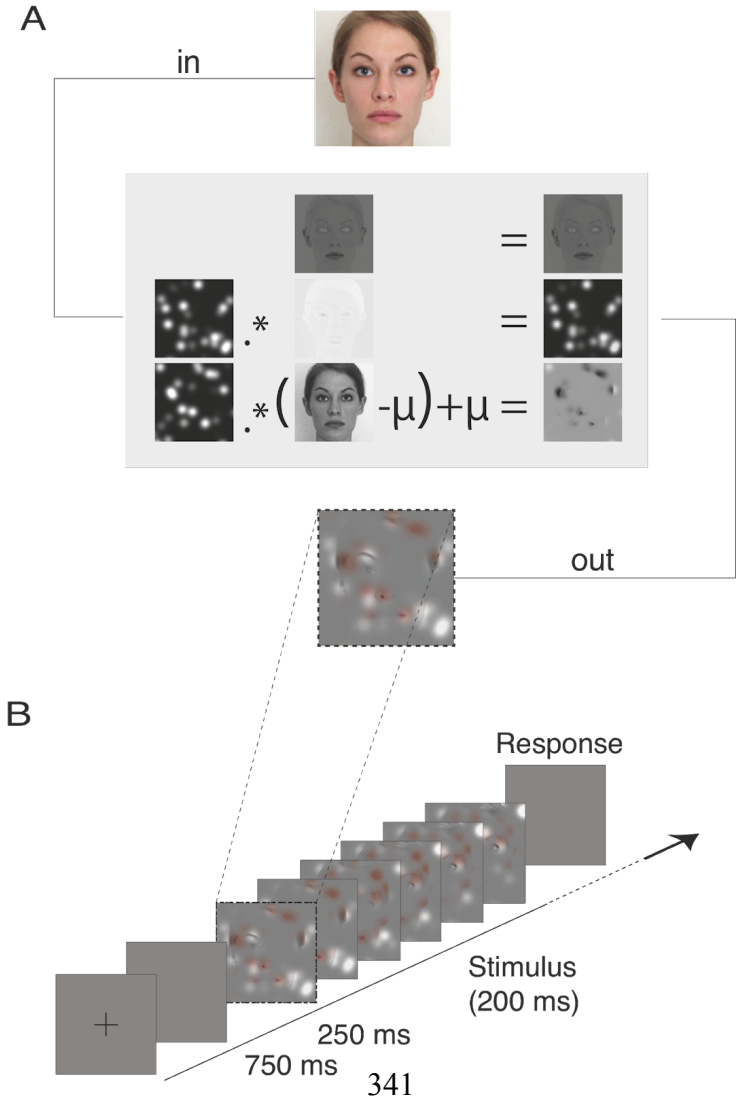
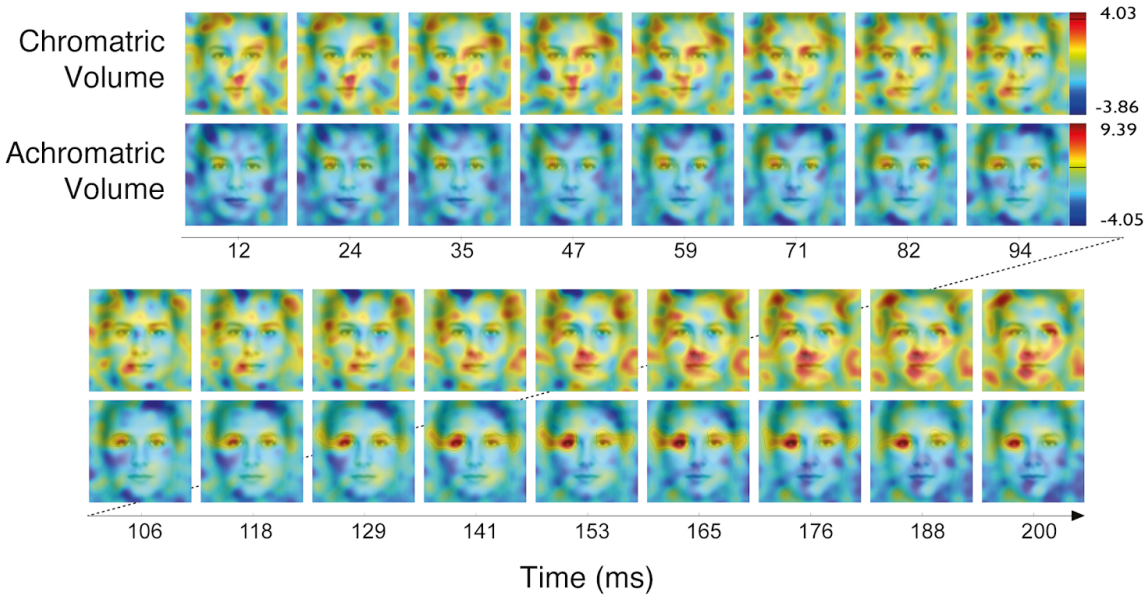


Figure 2. Linear classification volumes for facial sex categorization. A weighted sum of the volumes of chromatic Gaussian apertures (explanatory variables) by the participant's response accuracy (dependant variable) was computed. The same procedure was also done between the volumes of achromatic Gaussian apertures and the participant's response accuracy. Spatio-temporal regions containing significant regression coefficients for chromatic (2nd and 4th row) and chromatic face information (1st and 3rd row) are delimited by thin black lines (pixel-test, $p<.05$). The frames of the classification volumes were superimposed on the average face to help with interpretation.



Annexe III : Article 8

Stationary objects flashed periodically appear to move during smooth pursuit eye movement

Frédéric Gosselin & Simon Faghel-Soubeyrand

Département de psychologie, Université de Montréal

Correspondence concerning this article should be addressed to Frédéric Gosselin, Départment de psychologie, Université de Montréal, Montréal, Québec, Canada. E-mail: frederic.gosselin@umontreal.ca

Publié: Gosselin, F., & Faghel-Soubeyrand, S. (2017). Stationary Objects Flashed Periodically Appear to Move During Smooth Pursuit Eye Movement. *Perception*, *46*(7), 874–881. https://doi.org/10.1177/0301006617694188

Abstract

We discovered that a white disc flashed twice at the same location appears to move during smooth pursuit eye tracking in the direction opposite to that of the eye movement. We called this novel phenomenon *movement-induced apparent motion* or *MIAM* for short. Using the method of constant stimuli, we measured the required displacement of the second appearance of the disc in the pursuit direction to null the effect during the closed-loop stage of smooth pursuit eye tracking. We observed a strong linear relationship between the points of subjective stationarity and the inter-stimuli intervals for four smooth pursuit eye movement speeds. The slopes and y-intercepts of these linear fits were well predicted by the hypothesis according to which subjects saw illusory motion from the first to the second retinal projections of the flashed disc during smooth pursuit eye movement, with no extra-retinal signal compensation.

Introduction

When our eyes are still and two neighbouring visible lights are flashed in rapid succession, we see movement when nothing is actually moving in the environment. This phenomenon is named beta or apparent motion; it was first thoroughly studied by Wertheimer in 1912. A similar situation occurs on the retina when an observer engages in smooth pursuit eye movement (SPEM) in front of a stationary light flashed periodically. In both cases, neighbouring groups of retinal receptors are stimulated by light successively. Indeed, the two cases are indiscriminable *on the retina* if the following conditions are satisfied: i) if the light is first flashed at the same retinal location in both cases, ii) if the period of the flashes is the same in both cases, and iii) if the displacement vector of the lights in the apparent motion case divided by the period of the flashes is equal to the SPEM velocity in the other case.

As you would expect *qualitatively* from this retinal analysis, eye tracking the tip of a moving pencil or finger in front of Video 1 (see supplementary material) which displays stationary white discs flashed with periods increasing by steps of 16.67 ms from 0 ms, at the bottom, to 100 ms, at the top, results in the discs appearing to move in the direction opposite to that of the eyes. Importantly, *you need to pay attention to the flashing discs* while engaging in SPEM. As far as we know, this phenomenon—henceforth called *movement-induced apparent motion* or *MIAM* for short—has never been reported in the literature. You might have noticed during this demonstration that the white discs align with one another. This indicates that the apparent disc displacement is *proportional* to the speed of SPEM times the period of the flashes. Again, this is compatible with the above retinal analysis which predicts that the apparent disc displacement should be *equal* to the speed of SPEM times the period of the flashes.

However, it is at odds with our ubiquitous experience that the *continuously illuminated* world around us is stable despite shifts of the images projected onto the retina during SPEM. Helmholtz (1925) suggested that this retinal displacement is cancelled by an "effort of the will". Sperry (1950) developed the idea and proposed that the retinal displacement is compared to

the *corollary discharge* (also called *efference copy*; von Holst and Mittelstaedt, 1950)—a copy of the motor command that is sent to the muscles to produce an eye movement. Although our visual system comes close to this ideal, slight undercompensation can be revealed in the form of the perception of a tiny and usually non-disturbing movement of the stationary world opposite to eye movement. This illusionary motion of a *continuously illuminated* background induced by SPEM is referred to as the Filehne illusion (Filehne, 1920; Haarmeier, Thier, Repnow & Petersen, 1997). Given the obvious similarities between MIAM and the Filehne illusion, it is tempting to argue that they are one and the same. However, the small undercompensation for retinal slip, which is believed to cause the Filehne illusion, should be identical for all the pulsating white discs in Video 1; therefore, the Filehne illusion cannot explain why the discs appear to move less and less from top to bottom. Moreover, the magnitude of the erroneous motion characterising the Filehne illusion (e.g., that of the disc at the bottom of Video 1, which is presented continuously) is minuscule in comparison to that characterising MIAM (e.g., that of the disc at the top, which is flashed with a period of 100 ms). Jumping the guns, we shall show in the following pages that MIAM is very well explained by the SPEM analog of apparent motion described above, *with no extra-retinal compensation whatsoever*.

Specifically, we will take a *quantitative* look at the atomic event in Video 1: two discs flashed during SPEM. Using the method of constant stimuli, we will find the displacement of the second disc in the pursuit direction required to null MIAM. We will do this for several inter-stimuli intervals (ISI) and SPEM speeds. Complete failure to compensate for the large retinal slip between the appearance of the two white discs would lead to a linear relationship between ISI and the nulling displacement with a slope equal to the tracking speed and a y-intercept equal to 0; and full compensation would lead to no nulling required at all.

Methods

Subjects. Six observers (3 males) aged between 19-44 took part in the experiment. All participants had a normal or corrected-to-normal visual acuity. Written consent was obtained for all participants, and the study was approved by the Université de Montréal ethics committee.

Apparatus. The experimental programs were run on a MacPro computer in the Matlab environment, using the Psychophysics (Brainard, 1997; Pelli, 1997) and EyeLink toolboxes (Cornelissen, Peters & Palmer, 2002). All stimuli were presented on an Asus VG278H at 60 Hz with a resolution of 1920 x 1080 pixels. Room lights were turned on so that natural background visual cues, such as the frame of the computer monitor, were available. A chin rest was used to maintain viewing distance at 66 cm. Eye movements were monitored at 250 Hz with an EyeLink II head-mounted eye-tracker (SR Research, Mississauga, Ontario). Only the dominant eye—as assessed by the Miles test (Miles, 1930)—was tracked, but viewing was binocular.

Procedure. A nine-point calibration was performed with the eye-tracker at the beginning of each experimental session, and a drift-correction was performed every five trials. On each trial, a stationary red dot with a diameter of 6.50 arcmin was shown for 400 ms on a black background either 0.95, 2.90, 3.87 or 9.69 deg to the left of the unmarked center of the computer monitor, depending on SPEM speed. Then the red dot began to move rightward at a constant speed of 1.95 deg/s (FG3, SFS3), 5.85 deg/s (FG2, SFS2, and WM), 7.80 deg/s (CAC, CBP, FG1, LN, and SFS1) or 19.50 deg/s (FG4 and SFS4). When the red dot reached the unmarked center of the computer monitor, it was superimposed on a centred white disc with a diameter of 0.43 deg presented for 16.67 ms. This was followed by an inter-stimuli interval (ISI) which varied between 16.67 ms and 133.33 ms, depending on the experimental session. A second white disc, identical to the first one, was then presented for 16.67 ms with a horizontal offset relative to the center of the computer monitor + SPEM speed [deg/sec] x ISI [sec] of either -0.22, -0.13, -0.04, 0.04, 0.13, or 0.22 deg. See Videos 2 and 3 (see supplementary material). Each subject completed one series of either four (for the 19.50 and 1.95 deg/s SPEM speeds) or of seven (for the 7.80 and 5.85 deg/s SPEM speeds) experimental sessions of 120 trials. ISI varied between 16.67 ms and

133.33 ms in 33.33 ms steps for the series of four experimental sessions, and in 16.67 ms steps for the series of seven experimental sessions

Subjects were asked to eye-track the red dot, to pay attention to the white disc, and to indicate the direction in which the white disc appeared to move using computer keyboard keys. Each participant completed about 10 practice trials before beginning the experimental sessions. Two subjects (FG5 and SFS5) did an additional experimental session with a SPEM speed of 19.50 deg/s and with an ISI of 133.33 ms, identical to the one described above, except that all white discs' positions were shifted 1 deg leftward.

Results

For every trials of a subject, gaze location traces were filtered with a 2^{nd} order low-pass Butterworth filter with a cutoff at 50 Hz (Lindner, Schwarz & llg, 2001). Then instantaneous eye velocities were computed by deriving the filtered gaze locations; and instantaneous accelerations were computed by deriving these velocities. Saccades were defined as accelerations greater than 700 deg/s^2. Intervals of 60 ms centered on saccades were removed from the velocity traces and replaced by linearly interpolated velocities. Mean velocity traces after saccade removal for every subjects and SPEM speeds are shown in Figure 1. Finally, closed-loop SPEM gain was obtained by dividing the mean, across trials, of these velocities between 500 ms (which corresponds to the appearance of the first white disc) and 650 ms after red dot motion onset by the red dot speed. Individual pursuit velocity gains are shown in Table 1.

The mean gain across subjects and SPEM speeds—0.89—indicates that pursuit quality was slightly under the ideal conditions. However, the most determinant factor for the hypothesis according to which subjects saw illusory motion from the first to the second retinal projections of the flashed disc during smooth pursuit eye movement, with no extra-retinal signal compensation,

is the mean horizontal displacement, *including saccades*, not pursuit velocity gain. We thus computed the mean horizontal eye displacement error relative to the red dot horizontal displacement between the two flashed white discs across trials for every subjects and SPEM speeds (Table 1). The mean of these errors was only 5.36%.

Table 1. Individual points of subjective stationarity (PSS) fitted parameters and other measures.

Subject	Red dot speed (deg/s)	Pursuit gain	Eye displacement error (%)	Linear fit slope (deg/s)	Linear fit y-intercept (deg)	R^2
SFS4	19.50	0.85	4.33	19.02	0.03	.99
FG4	–	0.83	3.22	19.19	0.02	.99
SFS5	–	0.80	0.49	–	–	–
FG5	–	0.78	6.48	–	–	–
CAC	7.80	0.89	6.79	7.31	0.05	.99
CBP	–	0.89	6.20	6.84	−0.02	.97
FG1	–	0.82	7.26	7.63	0.01	.99
LN	–	0.93	8.26	7.66	−0.03	.99
SFS1	–	0.86	8.94	7.73	−0.03	.99
FG2	5.85	0.96	6.27	5.64	0.01	.99
SFS2	–	0.92	3.45	5.82	−0.03	.99
WM	–	0.97	0.36	5.93	−0.02	.99
SFS3	1.95	0.86	0.72	2.04	−0.04	.99
FG3	–	0.93	6.00	1.75	0.00	.98

Then, for each subject, for each ISI and for each horizontal offset, we computed the proportion of trials the subject responded that the white dot moved rightward. Next, we fitted cumulative Gaussian distributions to the six proportions—one for each of the horizontal offset—associated with each of the four or seven ISI (R^2 ranged from 0.87 to 1.00, with a mean of 0.99 and a standard deviation of 0.02), and extracted seven points of subjective stationarity (PSS). PSS are represented as black circles in Figure 2 (or as blue asterisks for the extra leftward-shifted sessions, Figure 2, A).

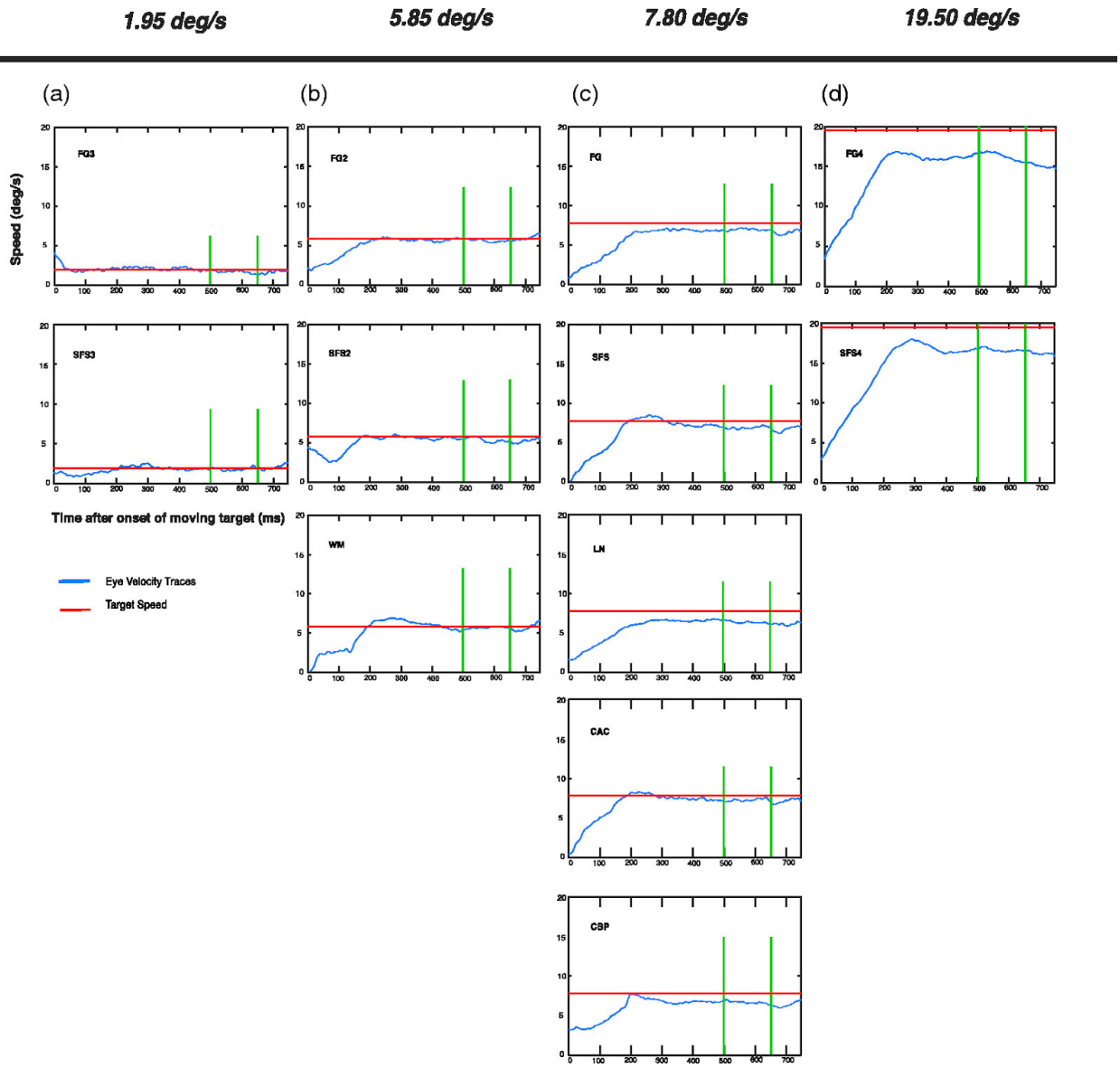


Figure 10.1. Blue curves represent the average individual filtered velocities as a function of time following onset of target motion after saccade removal. Red lines represent the speed of the moving targets. The green bars delimit the interval used to compute the gains (the first green bar coincides with the presentation of the first white disc).

Lines fitted very well to these PSS (Figure 2, black solid lines; Table 1, R^2), confirming the qualitative impression that the magnitude of MIAM is proportional to the ISI (or to the period in Video 1). Moreover, the slopes and y-intercepts of these linear fits were very close to those predicted by a SPEM analog of apparent motion, with no extra-retinal signal compensation whatsoever (Figure 2, solid red lines). The PSS of the two experimental sessions with a SPEM speed of 19.50 deg/s and an ISI of 133.33 ms—the one with the first white disc superimposed on the red dot and the other with the first white disc presented 1 deg to the left of the red dot (Figure 2, D)—are almost identical.

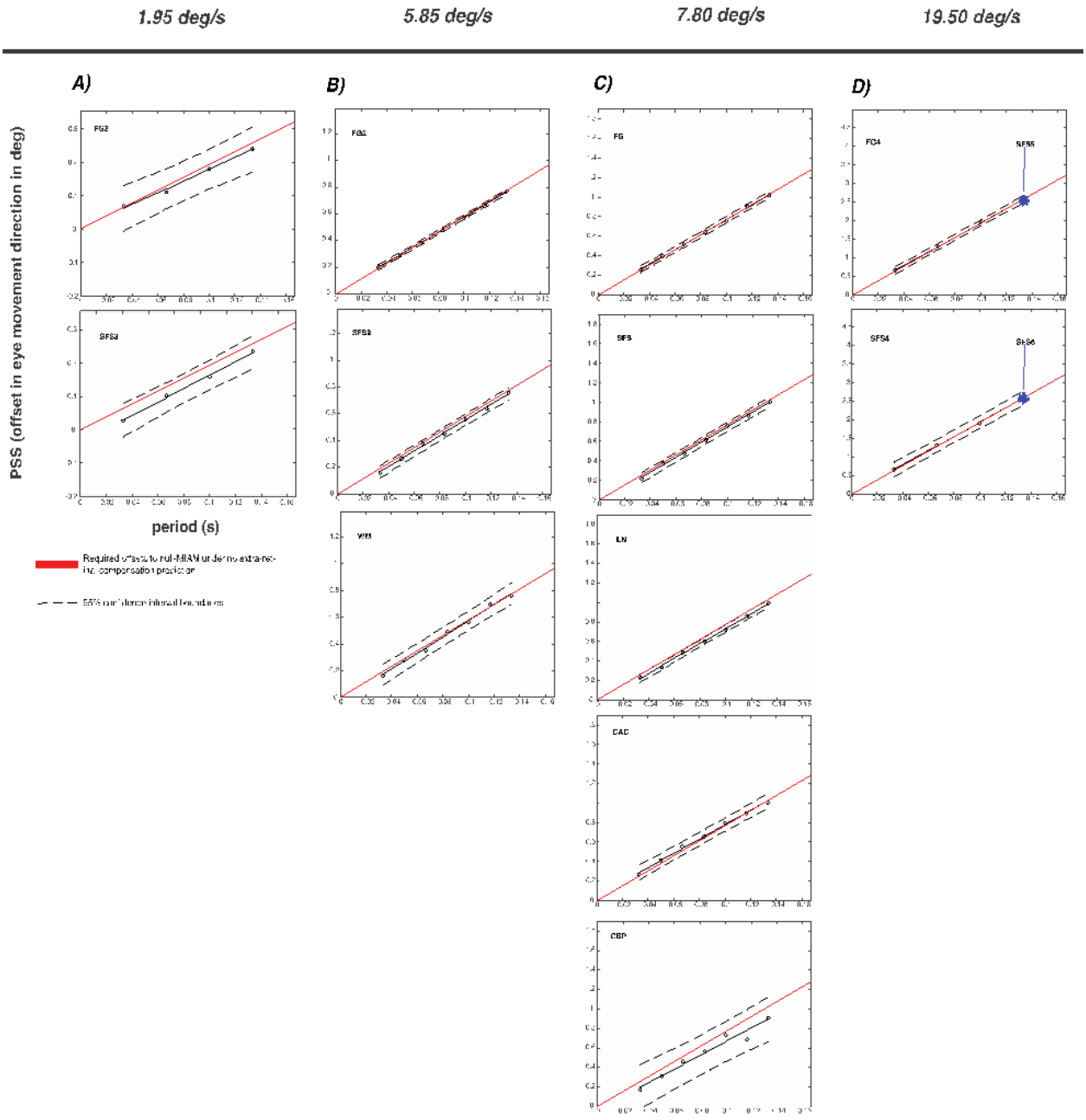


Figure 10.2. The magnitude of MIAM for every subject as a function of ISI and SPEM speed. Black circles represent points of subjective stationarity (i.e., the offsets of the second flashed white disc required to null MIAM); solid black lines are the best linear fit to these data points; dashed black lines indicate the boundaries of the 95% interval of confidence of the fitted lines; and the solid red lines is the prediction assuming a SPEM analog of apparent motion, with no

extra-retinal signal compensation whatsoever. Blue asterisks in the scatterplots are the PSS for the sessions during which SPEM speed was 19.50 deg/s, ISI was 133.33 ms, and the first white disc was presented 1 deg to the left of the red dot.

Discussion

We discovered that two identical white disc flashed at the same location appears to move during the closed-loop stage of smooth pursuit eye movement (SPEM) in the direction opposite to that of the eye movement. We called this novel phenomenon *movement-induced apparent motion* or *MIAM* for short. Using the method of constant stimuli, we measured the required displacement of the second disc in the pursuit direction to null the effect. We observed a strong linear relationship between the points of subjective stationarity and the inter-stimuli intervals for a given SPEM speed in six subjects. The slopes and y-intercepts of these linear fits were well predicted by the hypothesis according to which subjects saw illusory motion from the first and second discs' retinal projections during smooth pursuit eye movement, without extra-retinal signal compensation. This was the case for SPEM speeds between 1.95 and 19.50 deg/s, which indicates that MIAM is invariant to SPEM and apparent motion speed. This also suggests that SPEM and apparent motion share mechanisms in both slow and fast motion processes (see also Matsumiya & Shioiri, 2015). Finally, the points of subjective stationarity of the two experimental sessions with a SPEM speed of 19.50 deg/s and an ISI of 133.33 ms—the one with the first white disc superimposed on the red dot and the other with the first white disc presented 1 deg to the left of the red dot (Figure 2, D)—are almost identical, suggesting that MIAM is invariant to retinal location at least within the fovea.

Several studies converged to show that the middle temporal area in the macaque and other primates and its human homolog, the human MT complex, respond to stimulus conditions that induce apparent motion (Mikami, 1991;Goebel, Khorram-Sefat, Muckli, Hacker & Singer, 1998; Muckli et al., 2002; Liu, Slotnick & Yantis, 2004). The brain locus of the compensation for the retinal slip resulting from SPEM seems to be distributed along the visual pathway. Neuroimaging

studies in humans have failed to find any trace of a compensation for SPEM in area V1 up to area MT; the response in these visual areas seem to correlate with retinal motion signal, not with perceived motion (Haarmeier and Thier 1998; Lindner, Haarmeier, Erb, Grodd, & Thier, 2006; Trenner et al. 2008; Tikhonov, Haarmeier, Thier, Braun & Lutzenberger, 2004). However, there is evidence for SPEM compensation before area MT in the primate brain. Neurons encoding visual motion independently of SPEM—so-called real-motion cells—represent about 10-15% of the cells in areas V1 and V2 and about 40% of the cells in area V3A (Galletti & Fattori, 2003). Moreover, it was shown that about 8% of cells in area V1 exhibited percept related activity in the awake behaving monkey (Dicke, Chakraborty & Thier, 2008). With this in mind, one parsimonious neural explanation for the failure to compensate for the SPEM analog of apparent motion under MIAM conditions is that the critical compensation occurs either before the computation of apparent motion in the visual processing stream (in area V1, V2 or V3), or in a parallel pathway. In other words, the small retinal slip resulting from the brief presentation of the first disc would be cancelled along striate and early extra-striate visual areas, and so would the equally small retinal slip resulting from the brief presentation of the second disc but, most importantly, the much larger retinal slip resulting from the Gestalt grouping of the two discs in MT would not.

Haarmeier, Thier, Repnow & Petersen (1997) described the unique case of a patient with relatively large bilateral extrastriate cortex lesions, which experiences false perception of motion due to a complete inability to take eye movements into account when faced with self-induced retinal image slip. Probably as a result of the conflict this condition creates between the visual, vestibular, and somatosensory senses (Reason & Brand, 1975), this patient suffered from dizziness and nausea. What we have shown here is that for stroboscopically illuminated stationary objects such a complete incapacity of cancelling retinal displacement resulting from SPEM is the norm. We believe, therefore, that MIAM might play a role in *flicker sickness* in which observers report nausea and feeling dizzy when, for example, in a helicopter (Cushman & Floccare, 2007) or working with intermittently flashing lights (Ulett, 1953).

References

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

Cornelissen, F.W., Peters. E., Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments & Computers, 34*, 613-617.

Cushman, J. T., & Floccare, D. J. (2007). Flicker illness: an underrecognized but preventable complication of helicopter transport. *Prehospital emergency care*, *11*(1), 85-88.

Dash, S., Dicke, P. W., Chakraborty, S., Haarmeier, T., & Thier, P. (2009). Demonstration of an eye-movement-induced visual motion illusion (Filehne illusion) in Rhesus monkeys. *Journal of vision*,*9*(9), 5-5.

Dicke, P.W., Chakraborty, S. & Thier, P. (2008). Neuronal correlates of perceptual stability during eye movements. *European Journal of Neuroscience, 27*, 991-1002.

Erickson, R. G. & Thier, P. (1991). A neuronal correlate of spatial stability during periods of self-induced visual motion. *Experimental Brain Research, 86,* 608-616.

Filehne, W. (1922). Uber das optische wahrnehmen von bewegungen. *Zeitschrift fur Sinnephysiologie, 53*, 134-145.

Galletti, C., & Fattori, P. (2003). Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia*, *41*(13), 1717-1727.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H. & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent

motion and motion imagery. *European Journal of Neuroscience, 10*, 1563-1573.

Haarmeier, T., Thier, P., Repnow, M., & Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. *Nature*, *389*, 849-852.

Haarmeier, T., & Thier, P. (1998). An electrophysiological correlate of visual motion awareness in man. *Journal of Cognitive Neuroscience*, *10*(4), 464-471.

Haarmeier, T., & Kammer, T. (2010). Effect of TMS on oculomotor behavior but not perceptual stability during smooth pursuit eye movements. *Cerebral Cortex*, *20*, 2234-2243.

Helmholtz, H. V. (1925). Handbook of physiological optics. [Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Voss.]

Ilg, U. J. & Thier, P. (1996). Inability of rhesus monkey area V1 to discriminate between self-induced and externally induced retinal image slip. *European Journal of Neuroscience, 8*, 1156-1166.

Liu T., Slotnick, S.D. & Yantis, S. (2004). Human MT mediates perceptual filling-in during apparent motion. *NeuroImage, 21*, 1772-1780.

Lindner, A., Schwarz, U., & Ilg, U. J. (2001). Cancellation of self-induced retinal image motion during smooth pursuit eye movements. *Vision research*, *41*(13), 1685-1694.

Lindner, A., Haarmeier, T., Erb, M., Grodd, W., & Thier, P. (2006). Cerebrocerebellar circuits for the perceptual cancellation of eye-movement-induced retinal image motion. *Journal of Cognitive Neuroscience*, *18*(11), 1899-1912.

Matsumiya, K., & Shioiri, S. (2015). Smooth pursuit eye movements and motion perception share motion signals in slow and fast motion mechanisms. Journal of Vision, 15(11):12, 1–15, doi:10.1167/15.11.12.

Metzger, W. (1932). Versuch einer gemeinsamen theorie der pha ̈nomene fro ̈hlichs und hazelhoffs und kritik ihrer verfahren zur messung der empfindungszeit. *Psychologische Forschung*, 16, 176 – 200.

Mikami, A. (1991) Direction selective neurons respond to short-range and long-range apparent motion stimuli in macaque visual area MT. *International Journal of Neuroscience, 61*, 101-112.

Miles, W. R. (1930). Ocular dominance in human adults. *The journal of general psychology*, *3*(3), 412-430.

Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F.E., Singer, W, & Goebel, R. (2002). Apparent motion: Event-related functional magnetic resonance imaging of perceptual switches and states. *Journal of Neuroscience, 22*, RC219.

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

Reason, J.T. & Brand, J.J. (1975). *Motion Sickness*. NewYork, NY: Academic Press.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*, 482.

Tikhonov, A., Haarmeier, T., Thier, P., Braun, C., & Lutzenberger, W. (2004). Neuromagnetic activity in medial parietooccipital cortex reflects the perception of visual motion during eye

movements.*Neuroimage*, *21*(2), 593-600.

Trenner, M.U., Fahle, M., Fasold, O., Heekeren, H.R., Villringer A., & Wenzel, R. (2008). Human cortical areas involved in sustaining perceptual stability during smooth pursuit eye movements. Human Brain Mapping, 29, 300-311.

Ulett, G. A. (1953). Flicker sickness. *AMA archives of ophthalmology, 50*, 685-687.

Holst von E, M. H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, *37*, 464-76.

Van Beers, R. J., Wolpert, D. M., & Haggard, P. (2001). Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *Journal of Neurophysiology*, *85*(5), 1914-1922.

Wertheimer, M. (2012). Experimental studies on seeing motion. *On perceived motion and figural organization*, 1-91. [Wertheimer, M. (1912). *Experimentelle studien über das sehen von bewegung*. JA Barth.]

Supplementary Material

All videos can be found here (must run in loop):
https://journals.sagepub.com/doi/suppl/10.1177/0301006617694188

Video 1. In this video, the white discs are repeatedly visible for 16.67 ms and invisible for an inter-stimuli interval that varies between 0 ms for the disc at the bottom to 100 ms for the disc at the top, in successive steps of 16.67 ms. With the video playing in loop in the background, eye tracking the tip of a moving pencil or finger while paying attention to the flashing discs should result in the discs appearing to move in the direction opposite to that of the eyes, less and less from top to bottom. Note that the discs seem to align with one another. We call this novel phenomenon *movement-induced apparent motion* or MIAM, for short.

Video 2. Sample trial illustrating the basic MIAM phenomenon. A stationary red dot with a diameter of 8 pixels is shown for 400 ms on a black background to the left of the unmarked center of the video. Then the red dot begins to move rightward at a constant speed of 240 pixels/s. When the red dot reaches the unmarked center of the video, it is superimposed on a white disc with a diameter of 20 pixels presented for 16.67 ms. After an ISI of 100 ms, a second white disc, identical to the first one, is presented for 16.67 ms at the center of the video. Eye tracking the red dot while paying attention to the white disc should result in the disc appearing to move leftward.

Video 3. Sample trial illustrating the nulling of the basic MIAM phenomenon. This video is identical to Video 2 except for the following: The second white disc is presented with a horizontal offset of 24 pixels relative to the center of the video (i.e., 240 pixels/s * 0.10 s, as predicted under the assumption of a complete failure to compensate for the retinal slip caused by

smooth pursuit eye movement). Eye tracking the red dot while paying attention to the white disc should result in the disc appearing stationary.