



HAL
open science

The development of visual object categorization

Céline Spriet

► **To cite this version:**

Céline Spriet. The development of visual object categorization. Neuroscience. Université Claude Bernard - Lyon I, 2024. English. NNT : 2024LYO10231 . tel-04806717

HAL Id: tel-04806717

<https://theses.hal.science/tel-04806717v1>

Submitted on 27 Nov 2024

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

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

**THESE de DOCTORAT DE
L'UNIVERSITE CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 476
Ecole Doctorale Neurosciences et Cognition**

**Discipline :
Neurosciences**

Soutenue publiquement le 25/10/2024, par :
Céline Spriet

**The development of visual object
categorization**

Devant le jury composé de :

Macaluso, Emiliano	Professeur UCBL	Président
Höhl, Stefanie	Professeur University of Vienna	Rapporteuse
Op de Beeck, Hans	Professeur KU Leuven	Rapporteur
Dehaene-Lambertz, Ghislaine	Directrice de recherche Université Paris-Saclay	Examinatrice
Hochmann, Jean-Rémy	Directeur de recherche UCBL	Directeur de thèse
Papeo, Liuba	Directrice de recherche UCBL	Co-directrice de thèse
Gomez, Alice	Maître de conférences UCBL	Invitée



Acknowledgments

I am very grateful to my advisers, Liuba Papeo and Jean-Rémy Hochmann, for their unwavering support and amazing scientific mind. They never failed to support me along the way, so I wasn't alone during this journey. Thank you to Liuba for believing in me and supporting me from the start of my adventure; and to Jean-Rémy for teaching me so much and trusting me. I really appreciate you pushed me to go and share my work at several conferences; you know how difficult it was for me at first. Thank you both for giving me the possibility to be a scientist.

I would also like to give a special thank to my jury committee members: Ghislaine Dehaene-Lambertz, Alice Gomez, Stefanie Hoehl, Emiliano Macaluso, and Hans Op de Beeck, as well as James Bonaiuto and Marcela Peña, for their valuable time and evaluation of my work, allowing me to complete this journey.

There are times that require technical assistance; I want to thank Johan Pacquit in particular, whose serenity is so comforting that you can never feel anxious. Many thank to Thomas and Marco as well.

Doing research is not free, and we need generous organization to help us pay the different aspects of research, through grant. For this, I want to acknowledge the European Research Council Starting Grant awarded to Liuba Papeo (Project: THEMPO, Grant Agreement 758473); the ECOS-Sud action nr. C20S02, awarded to Marcela Pena, Jean-Rémy Hochmann and Judit Gervain; the Fondation de France who awarded a grant for the projet AttentioPrema (00134704 / WB-2022-46104) to Jean-Rémy Hochmann; the Fyssen Foundation grant awarded to Liuba Papeo; and finally, the Fondation pour la Recherche Médicale (FRM) who granted me with the Fin de Thèse grant, allowing me to have 1 more year to finish my PhD (FDT202304016547).

As I said before, I was never alone in this journey. I think all of us need a team of outstanding scientists to accomplish good work. They not only aid in gaining perspective on your work and analysis, but they also bring happiness and laugh in the lab. I am very grateful to Etienne, Nicolas, Parvaneh, Emilie, Sofie, Sara, Violette, Manuel, and Farzad for that, as well as for all the assistance they provide. And for the many students and post-doc of the institute; all so nice and caring.

I would like to thank all the participants, and the parents who bring their infants to the lab; without them, I wouldn't have been able to accomplish anything!

Now there are strengths that you gain from special people, for just sharing a moment with them. I can never be grateful enough to my special people: Anthony, Laurie, Marine, Aurélien, Olivier, Nicolas, Corentin. For the love and support, for helping me be a better person.

Finally, a special thank to my family and friends, who sometimes even travel from far to just help me having one more participant (Manon, Benjamin, and of course, Emma!).

I hope you'll enjoy reading the story of my journey – it is only a beginning.

General Abstract

Visual object categorization is at the interface between *seeing* and *thinking*. Evidences from functional MRI (fMRI) studies have described an organization of the ventral stream by visual categories, especially between animate and inanimate entities, that decomposed into finer-grained distinctions. How does this specific organization come about in humans? This thesis presents a set of 4 studies addressing this issue. First, I will hypothesize that the first categories infants represent are constrained by these dimensions organizing object representation in the visual cortex (chapter 3) and study the role of brain maturation and experience in this representation (chapter 4). Then, I will investigate how the speed of presentation influences the animate/inanimate categorization in the first year of life and in adulthood (chapter 5), and what visual features act in this categorization in adults (chapter 6).

Results show that infants will first be attracted by visual features that are not eliciting categorization, such as the size of stimuli, before relying on category membership, representing first the animate and inanimate entities. We suggest that representing more and more visual features, and being faster in representing the visual features, allows growing up infants to represent more (finer-grained) categories, and that being exposed to objects guide infants' behavior towards categories they already represent. Part of the big animate-inanimate categorization is based on low-level visual features (e.g., power spectrum) that covariate with the categories, but the behavior gets influenced by categories only when enough, more mid-level features are presented in the image.

Keywords: Visual categorization, infants' development, eye-tracking, EEG

General summary

When perceiving a visual stimulus, we immediately retrieve its category. Is it an animal, or an inanimate object? Is it a mammal, or a bird? Categorize objects surrounding us allows navigating in our environment and interacting with each encountered object. For instance, we can recognize a bench in a glance, and decide to sit on it, in order to rest. Categories are the first stage of the sensorial inputs interpretation, reflecting the visual features integration in a unique percept. The visual features, such as the color, the texture or the shape, are initially encoded into the visual cortex, all along the ventral stream. Visual categories will guide both our behavior and reasoning regarding the environment. Thus, visual categorization is at the interface between “*seeing*” and “*thinking*”.

The adult’s visual cortex activation is actually revealing an organization by visual categories (Haxby et al., 2011; Konkle & Caramazza, 2013; Op de Beeck et al., 2008, 2019). Indeed, studies have shown a big distinction between animate and inanimate object in the ventral stream, that crumble into finer-grained distinctions between human and nonhuman animal faces and bodies, and natural and artificial big and small inanimate objects. Remarkably, this organization is shared between species, as it was also evidenced in another nonhuman primate species, the macaque (Kriegeskorte, Mur, Ruff, et al., 2008). Moreover, it was evidenced in the visual cortex of congenitally blind individuals, while listening to sounds – either sounds naturally produced by objects, or objects’ name (Mahon et al., 2009; Mattioni et al., 2020; Striem-Amit et al., 2012; Striem-Amit & Amedi, 2014; van den Hurk et al., 2017). Altogether, it suggests that the visual object categories and its subsequent organization in the ventral stream of adults is biologically predisposed, and thus, might be present early in life. But it is not clear as to when those categories impact human’s behavior, how the brain maturation or visual experience influence the visual categorization, what is the speed of processing of those categories across development, or which visual features play a role in the visual categorization? This thesis aims at studying the development of visual object categorization: how and when visual categories emerge, influencing human behavior?

In chapter 3, I studied the visual object categorization in infants’ visual exploration using eye-tracking. I have shown that infants, as soon as 4 months of age, can represent images according to their categories: whether it is an animate or an inanimate object. However, they first explored images according to the size of the image on the screen, looking longer at the larger, more compact and less elongated image. Around 10 months, they lost their reliance on non-categorical visual features (i.e., features that are not eliciting the representation of a category) towards categorical ones (i.e., features

eliciting the representation of a category, as co-varying with it). They directly looked similarly at two images belonging to the same category, such as a cat and a fish, or a pear and a hat, but longer at animate compare to inanimate objects when seen next to each other. Growing up, 19-month-old infants represented more visual categories, especially the human and nonhuman categories, as well as categories of human body, animal face, animal body and natural small objects. Comparing infants' behavior with the adults' ventral stream activity measured by functional MRI (fMRI), I have also shown that while infants grow up and represent more and more visual categories, their behavior compares to larger part of the adults' ventral stream. This is suggesting that integrating more and more visual features, encoded in different portion of the ventral stream, allows to refine the representation of objects, thus represent more and finer-grained categories.

Chapter 4 follow the previous chapter, extending results to two new age groups, 6- and 8-month-old infants, as well as to a new population, preterm infants. Results reveals that at 8 months of age, infants explored images according to both categorical and non-categorical visual features. Before, at 4 and 6 months, infants only explored images according to their size, compactness and elongation, although they were able to represent the animate and inanimate categories, but only when the images were presented at the same size. After, at 10 and 19 months, infants only explored images according to their categorical features. At 8 months, infants did both, showing that one doesn't preclude the other, and that the switch observed between 4 and 10 months of age is a two-steps development; first, categorical features attract infants' look, then, non-categorical features lost interest and did not drive infants' attention anymore. Preterm results revealed that the first step of this transition highly depend on the experience; once infants represent a category, being exposed to objects will drive their attention towards the category.

Chapter 5 studied the impact of the speed of images presentation on their categorization into animate and inanimate objects, using EEG, in both adults and infants of 4 and 9 months. The stimuli set use in this study was, so far, to my knowledge, the larger stimuli set used to study the animate-inanimate categorization in 4- and 9-month-old infants. It was selected to be as much as possible representative of the tremendous variability observed in the world surrounding us, both between and within the animate and inanimate categories. This chapter shows an acceleration of the animate-inanimate capacity throughout life. Four-month-old infants couldn't categorize images when they were presented at about 167 milliseconds (ms), while 9-month-old infants could still categorize the same images when presented at about 83 ms, and adults, at 33 ms. Results show that, for each age, the faster the presentation of images, the lowest the categorization, as captured in the EEG signal. Thus, the speed

of images presentation does impact humans' ability to categorize objects, revealing the limits of the visual system when encoding visual categories. Such limitation of the visual system decreases by 3 between 4 and 9 months of age, and by 7 between 4-month-old infants and adults.

Finally, chapter 6 revealed that the big animate-inanimate categorization can be captured even based solely on low-level visual features, encoded early in the visual cortex of human adults. However, this categorization, captured by electroencephalography (EEG) in adults, do not impact adults' behavior. Indeed, they cannot name an image representing only the lower-level visual features of the original image, nor can they decide whether this image was originally an "animal" or not. However, it shows that the animate-inanimate categorization is resilient to the loss of visual information and that lower-level visual features only are sufficient to elicit such categorization in the adults' brain. It also shows that the animate-inanimate categorization captured in the adults' brain with intact images of the world cannot be explained by only few lower-level visual features.

Together, this thesis suggests that visual categories in human start, as soon as 4 months of age, by the big animate-inanimate categorization, representing objects based on higher-level visual features extracted in more anterior regions of the ventral stream. First, categorization will be hindered by reliance on lower-level visual features, before completely directing infants' visual exploration, once infants get sufficiently exposed to objects in their environment. Then, the ability to represent more and more visual features might allow infants to better represent images into finer-grained distinctions, and as infants grow older, more visual categories are represented (chapters 3 and 4). Infants need more time to be able to categorize object as animate or inanimate as compare to adults, but this capacity accelerates by an incredible amount between 4 and 9 months of age (chapter 5). Finally, adults' categorization show that the animate-inanimate categorization is feasible based only on lower-level visual features, but that those lower-level features alone cannot explain the whole animate-inanimate categorization, and are not sufficient to influence their behavior (chapter 6). It appears that categories (especially the animate and inanimate) will first be represented by extracting higher-level visual features from the objects, tuning lower-level areas that extract lower-level visual features towards features co-varying with the categories – at the same time, while infants grow older, they become faster in extracting features and integrating them together, allowing infants to represent objects faster. They also extract more and more visual features, allowing a more furnished representation of the objects, thus, more and finer-grained categorization of the objects.

Résumé général

Lorsque nous percevons un stimulus visuel, nous identifions immédiatement sa catégorie. Est-ce un animal ou un objet inanimé ? Un mammifère ou un oiseau ? Catégoriser les objets nous entourant nous permet de naviguer dans notre environnement et d'interagir avec chaque objet rencontré. Par exemple, on peut reconnaître en un coup d'œil un banc dans une rue, et décider de s'y assoir afin de s'y reposer. Les catégories constituent la première étape de l'interprétation de l'input sensorielle, reflétant l'intégration en un percept unique des propriétés visuelles, telle que la couleur, la texture ou encore la forme, initialement codées dans la voie ventrale du cortex visuel. Les catégories visuelles influencent notre comportement et notre raisonnement, qui sera ainsi adapté à notre environnement. Ainsi, la catégorisation visuelle se situe à l'intersection entre « *voir* » et « *penser* ».

Le cortex visuel des adultes est en fait organisé en fonction de ces catégories visuelles (Haxby et al., 2011; Konkle & Caramazza, 2013; Op de Beeck et al., 2008, 2019). On y trouve une distinction entre objets animés et objets inanimés, qui se décompose en plusieurs catégories plus détaillées, notamment entre visages et corps humains ou animaux (non-humains), ou entre les différentes tailles des objets (grands ou petits) artificiels ou naturels. Cette organisation spécifique est remarquablement partagée entre différentes espèces, notamment retrouvée chez d'autres primates non-humain, tel que les macaques (Kriegeskorte, Mur, Ruff, et al., 2008). Elle est également retrouvée chez des individus aveugles de naissance, lorsque ceux-ci écoutent des sons – soit naturellement produits par différents objets, soit directement le nom de ces objets (Mahon et al., 2009; Mattioni et al., 2020; Striem-Amit et al., 2012; Striem-Amit & Amedi, 2014; van den Hurk et al., 2017). Tout cela suggère que cette organisation serait prédisposée biologiquement, et contrainte à se développer. Ainsi, ces catégories pourraient être trouvées très tôt dans la vie d'un humain. Néanmoins, il n'est pas clair quand ces catégories émergeraient dans le comportement humain, comment la maturation cérébrale et l'expérience visuelle influenceraient cette émergence, quelle serait la vitesse d'intégration des catégories, ou quelles caractéristiques visuelles joueraient un rôle dans cette représentation des catégories ? Cette thèse étudie donc le développement de la catégorisation visuelle des objets : comment et quand ces catégories visuelles se mettent en place, influençant le comportement humain ?

Tout au long du chapitre 3, au moyen de techniques d'oculométrie, j'ai étudié et reporté le développement des catégories visuelles durant la première année de vie. J'ai notamment pu montrer que les nourrissons, à partir de 4 mois, peuvent représenter des images en fonction de leur catégorie, suivant qu'elles représentent un être animé ou un objet inanimé. Néanmoins, ils sont d'abord attirés par

la taille de l'image sur l'écran. Par la suite, vers 10 mois, les nourrissons perdent toute attirance envers la taille des images, et regardent aussi longtemps deux images appartenant à la même catégorie, par exemple un chat et un poisson ou bien une poire et un chapeau. En grandissant, vers 19 mois, les nourrissons représentent de plus en plus de catégorie, notamment les humains et les animaux, les corps humains, les corps animaux, les visages animaux et les petits objets naturels. En corrélant leur comportement avec l'activité mesurée tout au long de la voie ventrale d'adulte, par IRM fonctionnelle (IRMf), j'ai également mis en évidence qu'au plus les nourrissons grandissent et représentent de catégories, au plus leur comportement reflète l'organisation présente dans différentes parties du cortex visuel adulte. Cela suggère qu'être capable d'intégrer de plus en plus de caractéristiques visuelles, toutes codées à différents endroits dans la voie ventrale du cortex visuel, permet de représenter de plus en plus finement les images, en des catégories plus précises.

Le chapitre 4 complète le chapitre précédent, étendant ces résultats à des nourrissons âgés de 6 et 8 mois, ainsi qu'à des nourrissons nés prématurément. Il met notamment en évidence qu'à 8 mois, les nourrissons regardent les images non seulement en fonction de leur catégorie, mais également en fonction de leur taille, compacité et élongation. Cela révèle une transition en deux étapes : dans un premier temps, les caractéristiques catégorielles attirent l'attention des nourrissons, puis les caractéristiques non-catégorielles perdent leur importance et n'attirent plus leur regard. Les résultats des nourrissons de 8 mois nés prématurément montrent qu'ils regardent les images en fonction de leur taille, compacité et élongation, mais également en fonction des catégories « animé » et « inanimé ». Cela suggère que cette transition se produit lorsque le nourrisson a été suffisamment exposé à des objets.

Le chapitre 5 étudie l'impact de la vitesse d'apparition des images sur la catégorisation d'images entre objets animés et inanimés, en utilisant l'EEG, chez des adultes et des nourrissons âgés de 4 et 9 mois. Le set d'image a été sélectionné pour représenter, au maximum, la diversité de ces deux catégories dans le monde. C'est, à ma connaissance, jusqu'à présent, le plus grand set d'image utilisé pour étudier cette catégorisation chez des nourrissons. Ce chapitre met en évidence une accélération au cours du développement. Les nourrissons de 4 mois ne peuvent plus catégoriser les images si celles-ci sont présentées autour de 167 millisecondes (ms), alors que les nourrissons de 9 mois sont encore capables de le faire avec des images présentées pendant 83 ms, et les adultes, autour de 33 ms. Ainsi, la vitesse de présentation des images influe sur la capacité de catégorisation des humains, reflétant la limitation du système visuel lors de l'encodage des caractéristiques visuelles ; cette limitation se réduit déjà par un facteur 3 entre 4 et 9 mois, et réduit d'un facteur d'environ 7 entre 4 mois et l'âge adulte.

Enfin, le chapitre 6 révèle que la grande catégorisation entre objets animés et objets inanimés est en réalité déjà visible uniquement sur la base de caractéristiques visuelles codées très tôt dans le cortex visuels (caractéristiques visuelles de bas niveau). Néanmoins, cette catégorisation, mesurable par électroencéphalographie (EEG) chez l'adulte, n'influence pas leur comportement. En effet, les adultes seront incapables de nommer une image tellement déformée qu'elle ne contient plus que ces caractéristiques visuelles de bas niveau, ni même de dire si cette image appartient à la catégorie des « animaux » ou non. Toutefois, cela montre que chez les adultes, les caractéristiques visuelles de bas niveau permettent une catégorisation des objets visuels entre objets animés ou inanimés. Cette étude montre également que si ces caractéristiques visuelles suffisent à capturer une réponse catégorielle, elles n'expliquent pas entièrement la réponse observée lorsque l'image contient beaucoup plus de caractéristiques.

Ensemble, ces différentes études suggèrent que les catégories visuelles formées au cours du développement commencent, dès l'âge de 4 mois, par la représentation des objets animés et inanimés, basée sur la représentation de caractéristiques de haut niveau extraites dans des régions antérieures de la voie visuelle ventrale. Tout d'abord, cette catégorisation est camouflée par une attirance plus importante des caractéristiques visuelles de bas niveau, avant de complètement régir l'exploration visuelle des nourrissons. Puis, en étant capable de représenter de plus en plus de caractéristiques, les nourrissons sont alors capables de mieux représenter les images, en des catégories plus précises (chapitres 3 et 4). Les nourrissons ont besoin de temps pour représenter les catégories, mais le chapitre 5 montre une accélération conséquente au cours de la première année. Enfin, la catégorisation entre objet animés et inanimés est réalisable uniquement sur la base des caractéristiques visuelles de plus bas niveau (chapitre 6). Il semblerait que les catégories (notamment animé et inanimé) sont, dans un premier temps, représentées par l'extraction de caractéristiques de haut niveau, ce qui ajuste les aires représentant les caractéristiques de plus bas niveau vers les caractéristiques covariant avec les catégories – en même temps, alors que les nourrissons grandissent, ils deviennent de plus en plus rapide dans l'extraction et l'intégration des caractéristiques visuelles, ce qui leur permet de représenter les objets plus rapidement. Ils extraient également de plus en plus de caractéristiques visuelles, ce qui leur permet une représentation des objets plus fournies, et donc, de représenter les objets en plus de catégories, plus fines.

Contents

Acknowledgments.....	1
General Abstract.....	3
General summary	4
Résumé général	7
1. Introduction.....	12
1.1. Categorization at the basis of knowledge	12
1.2. The visual ventral stream.....	15
1.3. Visual categorization in infancy	19
1.3.1. Infants' vision.....	19
1.3.2. Behavioral evidence of early categorization	21
1.3.3. Neural evidence of early categorization	23
1.3.4. Summary of the early visual categorization	27
1.4. Thesis overview	29
1.4.1. Current research	29
1.4.2. Overview of chapters	29
2. General methods	31
2.1. Overview of the stimuli	31
2.1.1. Stimuli used in Chapter 3 and 4	31
2.1.2. Stimuli used in Chapter 5 and 6	31
2.2. The looking time in infancy.....	32
2.2.1. The looking time as a measure of dissimilarity	32
2.2.2. The Representational Similarity Analyses	33
2.3. Tagging a category-selective response in the EEG signal	34
2.3.1. Basics of electroencephalography (EEG).....	34
2.3.2. The frequency-tagging paradigm	35
3. The visual object categorization in infancy (study 1)	36
3.1. Introduction	36
3.2. Results	41
3.2.1. Experiment 1	41
3.2.2. Experiment 2	50
3.3. Discussion.....	52
3.3.1. Within 4 mo.....	53
3.3.2. From 4 to 10 mo.....	54
3.3.3. From 10 to 19 mo	55
3.3.4. Conclusions.....	57
3.4. Materials and Methods	58
3.4.1. Eye-Tracking Study.....	58
3.4.2. fMRI Study on Adults.....	61
4. The roles of experience and maturation in the development of categorical saliency (study 2)	62
4.1. Introduction	62
4.2. Materials and Methods	64
4.2.1. Participants	64
4.2.2. Stimuli	65
4.2.3. Analyses	65
4.3. Results	65
4.3.1. Experiment 1	65
4.3.2. Experiment 2.....	70
4.3.3. Experiment 3	72
4.4. Discussion.....	74
5. Perceptual acceleration enhances visual object categorization (study 3).....	78
5.1. Introduction	78

5.2.	Materials and Methods	81
5.2.1.	Participants	81
5.2.2.	Stimuli	82
5.2.3.	Experimental procedure	82
5.2.4.	EEG	83
5.2.5.	Analysis	84
5.3.	Results	87
5.3.1.	Experiment 1	87
5.3.2.	Experiment 2	88
5.3.3.	Experiment 3	89
5.3.4.	Comparison at different stages of development	90
5.4.	Discussion.....	90
5.4.1.	The animate-inanimate categorization	91
5.4.2.	The consequences of slower categorization abilities.....	91
5.4.3.	Conclusion.....	92
6.	Animate-inanimate object categorization from minimal visual information in the human brain and deep neural networks	94
6.1.	Introduction	94
6.2.	Experiment 1	96
6.2.1.	Materials and Methods	96
6.2.2.	Results	100
6.3.	Experiment 2	104
6.3.1.	Materials and methods	104
6.3.2.	Results	104
6.4.	Experiment 3	105
6.4.1.	Materials and Methods	105
6.4.2.	Results	106
6.5.	Behavioral judgments	108
6.5.1.	Materials and Methods	108
6.5.2.	Results	109
6.6.	DNN data.....	110
6.6.1.	Materials and methods	110
6.6.2.	Results	112
6.7.	Discussion.....	115
7.	General Discussion	119
7.1.	Summary of findings	119
7.2.	Infants represent visual categories from 4 months of age.....	121
7.2.1.	Infants' gaze is guided by category membership	122
7.2.2.	The special case of the animate-inanimate categorization	124
7.3.	Factors influencing the animate-inanimate categorization	126
7.3.1.	The animacy categorization is recruiting many features	127
7.3.2.	Acceleration of visual object categorization throughout life.....	129
7.4.	Open questions and future directions	130
7.5.	Conclusion.....	132
8.	Bibliography	133
9.	Supplementary materials.....	158
9.1.	Chapter 3 – Supplementary results	158
9.2.	Chapter 4 – Supplementary results	166
9.3.	Chapter 6 – Supplementary results	179

1. Introduction

1.1. Categorization at the basis of knowledge

Imagine this situation: you are at home, lying on your couch, peacefully reading a book – or this thesis. Your kid is napping in their room. Your cat is meowing, and jumps next to your head: you pat him a little, and understand that he is hungry. You feed him with some biscuits. As you are now standing, you noticed the plants are a little dry; thus, you water them. And now your kid is awake, calling you. This scene describes what could be a typical day, right? Everything seems natural and easy. In this situation, we do not even really think about anything, we just act. Yet, in order to interact and behave accordingly (e.g., giving biscuits to the cat rather than the plants; watering the plants rather than the couch), we have to know and recognize all the objects¹ surrounding us, including the cat, the book, the plant, and so on. Typically, we come across much more objects than what described in this little scenario, even some that we never have seen before – and somehow, we still manage to make sense of every object we possibly encounter in our daily lives. For instance, we might encounter new costumers every day, and we effortlessly know that we can interact with each of them – or, we might be confronted to a nut cracker for the first time, and yet, we already know that it’s an inanimate object. We might not understand the use of it right away, but we still recover its basic properties – it will not move by itself and disappear, it will fall if we don’t hold it, it will not be able to pass through other objects, it’s definitely a tool, etc. – allowing us to act accordingly. To access and use our knowledge about objects relies on cognition (Bayne et al., 2019). One key characteristic of our object knowledge is that it is not random, but organized into groups, that constitute the unit of our thought and are called categories (Carey, 2009; Harnad, 2005; Lloyd & Rosch, 1978). Categorization is thus at the basis of our cognition, organizing our thoughts, and our interaction with our surroundings.

Categorization helps structuring our mind and making sense of our environment (Carey, 2009; Mareschal & Quinn, 2001; Rosch, 1978), even from a young age (Gelman & Markman, 1986; Gelman & Meyer, 2011; Harnad, 2005), as it is the first step in accessing the knowledge associated with objects. When encountering a new object, such as a new cup, we recognize features diagnostic of a category we already know about (e.g., cup), and thus, we recognize it as a member of this category. Categories are the reason why we are able to generalize knowledge of previously encountered category members to novel ones (Davis & Poldrack, 2014). As structuring our thoughts and knowledge, categories influence

¹ In this thesis, the term “object” refers to any material entity, including alive ones such as animals.

up until our language. The grammatical rules of most languages have indeed, as the most frequent structure for sentences, a subject that is most of the time a human or a non-human animal, followed by a verb that describes an action, and an object, animate or inanimate (Dahl, 2008). Neuropsychology studies have identified impairments in category ability, describing patients having difficulties not only in naming/recognizing some objects, but also in retrieving associated knowledge (Capitani et al., 2003; Caramazza & Mahon, 2003; Mahon & Caramazza, 2009; Tyler & Moss, 2001; Warrington & Shallice, 1984). Most of the time, those impairments are category-specific; meaning that it is a whole category (e.g., edible things) that is impacted, but not other categories. These category-specific impairments might affect patients' life, as for instance, if you do not recognize edible things anymore, you can try eating random objects, including ones that are not edible (Mahon & Caramazza, 2009; Warrington & Shallice, 1984). Categories are really the unit of our thoughts, at the basis of our inferences and decision making, organizing the knowledge in our brain.

However, there is a virtually infinite number of categories, and an object can be categorized in many of them at the same time (Anderson, 1991; Bruner et al., 2017; Rosch, 1978). Categorization is defined as the ability to recognize different objects as members of the same group, discriminating them from members of other groups. Different levels of categorization co-exist, from the most inclusive to the most exclusive ones. Take a cat, for instance; it can be categorized into the broad, superordinate category commonly called "animal"; into the more specific category of "mammal"; obviously into the basic-level category of "cat"; into the even more specific subordinate category of "ragdoll", etc. (Mervis & Rosch, 1981; Rosch, 1978; Rosch & Mervis, 1975). Three levels of categorization are commonly described, the superordinate, subordinate and basic-level categories, although many levels exist within superordinate and subordinate levels (e.g., mammals and animals are both superordinate categories). Superordinate and subordinate categories are both established from the basic-level categories. The basic-level category (e.g., cat, dog, chair, ball) is the level we are most likely thinking about, talk about and thus refer to in our everyday life. It is the category in which objects have the optimal common features (diagnostic of the category) but with still enough features allowing to separate it from another category: it maximizes the difference between-category and optimizes the similarity within-category. For instance, a cat and a dog are similar in many ways; but also, very dissimilar on many ways – allowing a good separation of the 2 categories. On the other hand, if you consider a poodle and a coquer spaniel (two specific dog breeds), they are very similar to each other, and only few differences allow to make the distinction between both, making less obvious categories. The superordinate categories (e.g., mammals, animals) regroup a lot of objects that have fewer common features. Those

categories are all larger categories compare to the basic-level, regrouping the basic-level categories into larger groups. For instance, dogs and cats are part of the same superordinate category, but will still be separated from an apple and a chair. Subordinate categories (e.g., ragdoll, poodle) regroup fewer objects that are even more alike than objects at the basic-level. They refine objects into smaller groups as compare to the basic-level categories. Indeed, not a lot of objects looks more like a maine Coon than a ragdoll, yet they are part of different subordinate categories. A hierarchy in our access to the different level seems to exists: human adults would, as soon as they see an object, recognize and name this object at its basic-level (Grill-Spector & Kanwisher, 2005; Rosch, 1978; Thorpe et al., 1996; VanRullen & Thorpe, 2001). We will not access all categories of an object at once, but probably the ones that are of most importance for our everyday life, or a given goal – thus, except maybe if we are huge fans of cats, we will not identify specifically a ragdoll and a maine coon, but just two cats, retrieving that they are both pets, animals that can move, potentially bite when unhappy or threatened, but otherwise quite inoffensive. As our learning evolves with age and experience, the different levels of categorization might actually evolve, with the categories infants and children access being broader and more inclusive, or less distinctive, than the adults’ basic-level categories (Behl-Chadha, 1996; Gelman & Meyer, 2011; Pauen, 2002). But what are the first categories that influences the organization of human cognition? How do categories emerge in human life? How do infants see the world? Do very young infants already categorize things in their environment, even when having little, if any experience at all with those? Those questions will be further addressed in this thesis. But first, to be able to start answering those questions, we must first understand how categorization emerges in the adults’ brain?

If for now I described categorization in general, what categories allow us to do, by organizing our knowledge with the concept attach to them, I will now focus on the first step of categorization, and the main topic of this thesis – the visual categorization. Before being able to retrieve any categories, we must first perceive, even for only few milliseconds, the object we will categorize – we can use our hands and explore the object by touching it, we can also taste and/or smell some objects, or experience their sound. But one of the main sense we use to explore objects, is vision. Every morning, when we wake up, we start our day by opening our eyes and turn on the light or open the curtains. Observing the things around us comes often before navigating and interacting with them. Early in life, infants are also very limited in their movements, and first visually explore their environment before being able to move inside and interact further with objects using other sense. Thus, it is only logical that our visual system is already tuned towards categories. Indeed, in addition to patients presenting category-specific impairments – and associated brain lesions (Capitani et al., 2003; Caramazza & Mahon, 2003; Mahon

& Caramazza, 2009; Tyler & Moss, 2001; Warrington & Shallice, 1984) clearly suggesting that the brain is functionally organized into categories, patients with visual agnosia or lesion in the visual cortex clearly suggest the existence of a specific organization already in the visual areas of the brain (Biran & Coslett, 2003; Farah, 2004). For instance, lesions to the visual cortex has been suggested to cause selective deficits in naming animals, but not body parts (Caramazza & Shelton, 1998). Those impairments of human adults' vision suggest that categories can be extracted already from visual inputs, not only on knowledge (semantic information) associated with objects (e.g., animals can move while inanimate objects cannot). Adults' brain, by processing visual features, extract the visual categories from the visual input. By studying the organization of visual inputs in the part of the visual cortex dedicated to object recognition (the visual ventral stream), one can identify the main dimensions driving object categorization in humans. Indeed, with its major role in object recognition, the visual ventral stream is described as being at the interface between perception and cognition, supporting semantic categorization and representation of object in the rest of the brain (Carlson, Simmons, et al., 2014; Mahon & Caramazza, 2009).

1.2. The visual ventral stream

The retina is composed of two main cell types, two photoreceptors respectively called rods and cones (Wandell, 1995). Those cells project to the lateral geniculate nucleus (LGN) (Watanabe & Rodieck, 1989) and, from here, to the primary visual cortex V1 in the occipital lobe (Chatterjee & Callaway, 2003). From the primary visual cortex, two main processing streams are usually described: a ventral stream and a dorsal stream (Goodale & Milner, 1992; Haxby et al., 1991; Mishkin et al., 1983; Ungerleider & Haxby, 1994). The dorsal stream is directed into the parietal lobe and is recruited for processing the spatial location of objects (Mishkin & Ungerleider, 1982). On the other side, the ventral stream is directed into the temporal lobe and has been described as essential for the visual object recognition (Grill-Spector & Weiner, 2014; Reddy & Kanwisher, 2006), and is thus of great interest in the understanding of visual object categorization. It is often called the “what” pathways, and is the place of encoding of objects from visual inputs, by the integration of the different features extracted all along the ventral stream into a unified percept (Biederman, 1987; Marr & Nishihara, 1978; Wagemans et al., 2012). Typically, a first hierarchical organization is described all along the ventral stream, where we find the early, intermediate and anterior visual areas, each encoding different level of abstraction of the visual features. Early visual areas typically extract low-level visual features, intermediate areas mid-level visual features and anterior areas high-level visual features. (Grill-Spector et al., 1998; See for

review Grill-Spector & Malach, 2004; Hubel & Wiesel, 1998; Malach et al., 1995). Low-level features correspond to local measurements such as orientation, spatial frequency, shape or color. As information travel through the ventral stream towards more anterior regions, intermediate areas encode mid-level visual features, which are more global than the low-level features, and correspond to the global form (global curvi-linearity) of the stimuli and the texture (Long et al., 2017; Long & Konkle, 2017). Finally, the high-level visual features correspond to the integration of lower-level features into a comprehensive part, into an entity, encoding the most global features such as the shape of the object (Biederman, 1987; Marr & Nishihara, 1978). Another organization is described in each visual areas: the visual field maps, also called retinotopic maps. Those maps relate to the preservation of the spatial arrangement of the features in an image, meaning that central information in the visual field will be encoded in central part of visual areas while objects at the periphery will be encoded in part at the periphery of visual areas (Grill-Spector & Malach, 2004; Wandell et al., 2007). This specific organization allows us to recognize an object even if some pixels are removed from the images. However, the recognition becomes far more difficult as soon as the spatial arrangement of the part of the objects (e.g., eyes above nose above mouth configuration of a face) is disturbed.

The ventral stream also reveals an organization by visual categories, as widely described over the past years, using functional magnetic resonance imaging (fMRI). Indeed, categories already come out from the topographical activity of the ventral stream for different visual objects. The main categories found by studying the ventral pathway topography are the big animate and inanimate objects: all along the ventral stream, in both hemispheres, the medial part shows higher activity for inanimate objects as compare to animate ones, while the lateral part of the ventral stream shows higher activity for animate objects compare to inanimate ones (Bell et al., 2009; Chao et al., 1999; Downing et al., 2006; Konkle & Caramazza, 2013; Long et al., 2018; Wiggett et al., 2009). This big topographical distinction into animate and inanimate categories further breaks up into finer-grained categories, of which human and nonhuman animals (Papeo et al., 2017), and small and big (in terms of real-world size) objects (Konkle & Caramazza, 2013; Konkle & Oliva, 2012b; Long et al., 2018; Magri et al., 2020). Underneath this organization lies a mosaic of *hot spots* with the highest selectivity for biologically relevant stimuli such as faces, bodies and scenes (Downing et al., 2001; Epstein & Kanwisher, 1998; Grill-Spector & Weiner, 2014; Kanwisher et al., 1997; McCarthy et al., 1997; Pitcher et al., 2009; see for reviews Bracci & Op De Beeck, 2023; Op de Beeck et al., 2008, 2019).

Categorical distinctions also emerge from the different pattern of neural activity evoked by each objects (Bobadilla-Suarez et al., 2020; M. A. Cohen et al., 2017; Davis & Poldrack, 2013, 2014; Haxby

et al., 2011; Kriegeskorte, Mur, Ruff, et al., 2008). A pattern of activity reflects how the different part (i.e., the different voxels, smallest region observed in fMRI studies) of a region (e.g., the ventro occipito-temporal cortex, located in the ventral visual pathways) react to an object. Strikingly, patterns of activity are more similar (i.e., same voxels activated and same voxels non-activated) one another for two animate objects (e.g., a cat and a fish) than between one animate object and one inanimate object (e.g., a cat and a nut cracker). Patterns of activation within the ventral stream reveal not only the animate and inanimate categories, but also the finer-grain distinction between human and nonhuman animals, between the big and small inanimate objects, between the animal faces and bodies, as well as between the artificial and natural inanimate objects (M. A. Cohen et al., 2017; Kriegeskorte, Mur, Ruff, et al., 2008).

Interestingly, those categories, represented in the ventral pathway in the visual cortex, appear to be relevant for human adults' behavior, accounting for the way individuals parse the visual world. Indeed, finding a target-object among a set of distractors seems to be faster when the target belongs to a distinct visual category than the distractors (e.g., finding a cat among artificial objects) as compare to when the target and the distractors belong to the same visual category (e.g., finding a cat among animals including mammals). The search time increases as the neural similarity between the target and the distractors increases (Carlson, Ritchie, et al., 2014; M. A. Cohen et al., 2017). In addition, the activity recorded in the high-level visual region corresponds to what the subject reported to perceive, more than what was physically present. Indeed, the category-selective activation in the high-level visual cortex occurs only when the subject reported to perceive the objects (Grill-Spector, 2003). Moreover, proof of an animate-inanimate categorical distinction and other categorization at more basic-levels in adults' behavior exist, even with a high speed of presentation (Bacon-Macé et al., 2005; Grill-Spector & Kanwisher, 2005; Konkle & Oliva, 2012a; Long & Konkle, 2017; VanRullen & Thorpe, 2001).

Categorization in the adults' brain was not only evidenced in fMRI study; electroencephalography (EEG) and magnetoencephalography (MEG) studies also reveal categorization, and its fast emergence in the brain signal. Indeed, it is possible to decode from the EEG activity the signal recorded for animate stimuli from the signal recorded for inanimate objects stimuli, already 80 ms after the onset of the stimuli presentation. Other categories are also decodable from both signal, including human and nonhuman faces and bodies, as well as natural and artificial inanimate objects (Carlson et al., 2013; Cichy et al., 2014; Coggan et al., 2016; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; Proklova et al., 2016, 2019; Stothart et al., 2017; Thorpe et al., 1996). A lot of research evidenced especially the human face

categories, as opposed to all other categories (Gentile & Rossion, 2014; Quek & De Heering, 2024; Quek & Rossion, 2017; Rekow, Baudouin, Durand, et al., 2022; Rossion, 2014a, 2014b; Rossion et al., 2015, 2020; Rossion & Boremanse, 2011). If the spatial resolution of EEG and MEG is not as good as fMRI (Pfister et al., 2014), data from one measure can be related to data from another measure (Kriegeskorte, Mur, & Bandettini, 2008). In particular, MEG data were shown to first, in the first hundreds of milliseconds following the stimuli presentation, correlate with the activity recorded in V1, before being more related to the inferior temporal cortex in more anterior part of the ventral pathway (Cichy et al., 2014), adding evidence to a categorical organization in the human adult ventral stream.

If categorization is already captured in V1 and early in time, it suggests that visual categorization can be already extracted from low- and mid-level visual features, even before the contribution of high-level visual features (Kramer et al., 2023). Indeed, if shape is an important feature in the visual object categorization (Ayzenberg & Lourenco, 2019; Bracci et al., 2019; Bracci & Op de Beeck, 2016; Chen et al., 2023; Schmidt et al., 2017; Tiedemann et al., 2022), it is not the only feature eliciting categorization, and does not account for the whole categorical representation in the ventral pathway (Bracci & Op De Beeck, 2023; Jozwik et al., 2022; Kaiser, Azzalini, et al., 2016; Kramer et al., 2023; Long et al., 2017, 2018; Magri et al., 2020; Proklova et al., 2016; Ritchie et al., 2020; Rosenthal et al., 2018; Wang et al., 2022; Zeman et al., 2020). However, it is not always clear what account for the visual categorization in the adult's brain, as demonstrated by all the studies looking for features explaining the broad animate-inanimate distinction, exploring the role of shape, texture, global form of the stimuli, presence of faces and bodies, dimensions such as “being alive” (Bracci et al., 2019; Chen et al., 2023; Jozwik et al., 2022; Long et al., 2017; Ritchie et al., 2021; Schmidt et al., 2017; Wang et al., 2022). All those studies, of which here are only a subset, found evidence for each features acting in the animate-inanimate distinction.

All along this section, I described the organization of the human adult ventral stream, sometimes mixing in studies on nonhuman primates. Indeed, the organization of information in the humans visual cortex appears to be a hallmark in the evolution of the primate brain: it is replicated in the monkeys' inferior temporal cortex (Bell et al., 2009; Kiani et al., 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Pinsk et al., 2005; Sigala & Logothetis, 2002; Tompa & Sáry, 2010) but also in other animals (See for review Freedman & Miller, 2008). An intriguing characteristic is also that this organization, in the visual cortex, is largely resistant to variations of individual visual experience. Indeed, congenitally blind individuals present a ventral stream organization similar to the one observed in sighted individuals when hearing object sound, or object names (Mahon et al., 2009; Mattioni et al.,

2020; Striem-Amit et al., 2012; Striem-Amit & Amedi, 2014; van den Hurk et al., 2017). These results – i.e. the shared organization between species and across large differences in visual experience – suggest a neural code optimized by evolution: one might believe that the dimensions that drive the organization of visual object information in the adults’ brain are biologically disposed and determined (Caramazza & Shelton, 1997), and thus, may be observable early in life.

1.3. Visual categorization in infancy

1.3.1. Infants’ vision

Before diving into the question of whether infants already categorize objects in their environment, one can wonder whether infants see at all. The visual system, starting from the eye itself, is indeed developing from birth to adulthood, revealing an immaturity of the system at birth and in infancy (Fielder et al., 1988). However, an immaturity of the visual system does not mean that the visual system is not functional at all; it might be just functional enough to allow newborns and infants to see what they need to see, and not be overwhelmed by the extremely high amount of information surrounding them from the moment they open their eyes (Hainline, 1998; James, 1890). Already at birth, newborns, when opening their eyes, can detect a white line on a black background (Lewis et al., 1978). They do not only see and discriminate lines, but look as well – they actively move their eyes and head to obtain information about the world, to follow objects for instance (Candy, 2019; Gibson, 1988). However, even if they see, their vision is not adult-like yet. First, there is a high prevalence of astigmatism – light beams are not focused on the retina, but instead are spreader than what they really are, thus rendering a blurred image on the retina – in infancy, higher than in adults (Howland et al., 1978; Mohindra et al., 1978). This astigmatism will disappear in most infants in the first or second year of life, and influence infants’ acuity. Previous studies have measured Vernier acuity and Grating acuity, finding that both acuities develop very fast in the first year of life (Braddick et al., 1986; Cavallini et al., 2002; Gwiazda et al., 1978, 1980, 1997; Morison & Slater, 1985; Salapatek et al., 1976; Shimojo et al., 1984; Skoczenski & Norcia, 2002; see for reviews Dobson & Teller, 1978; Gerhardstein et al., 2009; Kellman & Arterberry, 2007; Slater, 2002). Vernier acuity is the smallest offset between 2 segments (or pattern, e.g., squares) allowing the discrimination between the two and thus their discrimination from a unique, simple line; grating acuity is the maximum number of grating (parallel segments) that can be discriminate from a uniform patch (see Figure 1A). At 6 to 8 months of age, infants’ Vernier acuity is 6 to 8 times smaller than adults’ acuity, and about 3 times smaller than what

it will be during childhood and early adolescence. Grating acuity will double from 6 months onward and reach the adult-level at 6 years of age (Skoczenski & Norcia, 2002). Thus, infants see less details than adults, but still see (for a simulation of infants' visual acuity at birth, 4 months and 9 months compared to adults, see Figure 1B). They can, from birth, discriminate orientations of gratings (Atkinson et al., 1988; Slater et al., 1988). They also encode size and look longer for bigger objects as compared to smaller identical objects such as simple cubes (Slater et al., 1990). Later on, when they can move more and start grasping objects, 6-month-old infants will reach smaller, more graspable objects with their hands, despite still looking more at bigger object (C. Newman et al., 2001). Binocular vision develops only 2 months after birth, allowing infants to see depth (Jandó et al., 2012; see for review Norcia & Gerhard, 2015). They can then, at 3.5 months of age, follow, with their eyes, a stereoscopic form (3D form) moving in one direction or another (Fox et al., 1980). Infants can also discriminate colors such as blue, green, yellow, orange, red and purple from white as early as 2 months of age (M. Bornstein et al., 1976; M. H. Bornstein et al., 1976; Teller et al., 1978; Zemach et al., 2007), and even discriminate blue from green and purple at 4 months (Franklin & Davies, 2004; see for review Skelton et al., 2022). However, newborns might not be able yet to discriminate so many colors, and first discriminate red from other colors (Adams, 1995). Infants' color vision is yet different than adults one (for a simulation of infants' vision of color, see Figure 1B), as the discrimination of a color from the background depends on the saturation of color, and infants' threshold is higher than adults' one, meaning that they cannot see less saturated colors compare to adults' capacity (Knoblauch et al., 2001).

Infants are able to see basic features (i.e., color, depth, shape, pattern, size...), and can use those features to discriminate an object from another object (Ayzenberg & Lourenco, 2022; Wilcox, 1999; Wilcox et al., 2008). For instance, from 3-4 months of age, infants can discriminate geometrical shapes such as a triangle from a square (Bomba & Siqueland, 1983; Quinn, 1987). As they are able to discriminate objects from each other, based on their dissimilarities, they also see their similarities. It is thus possible that, by seeing similarities between objects, they form visual categories early on, and represents objects sharing similar features more similarly as compare to objects having less similar features.

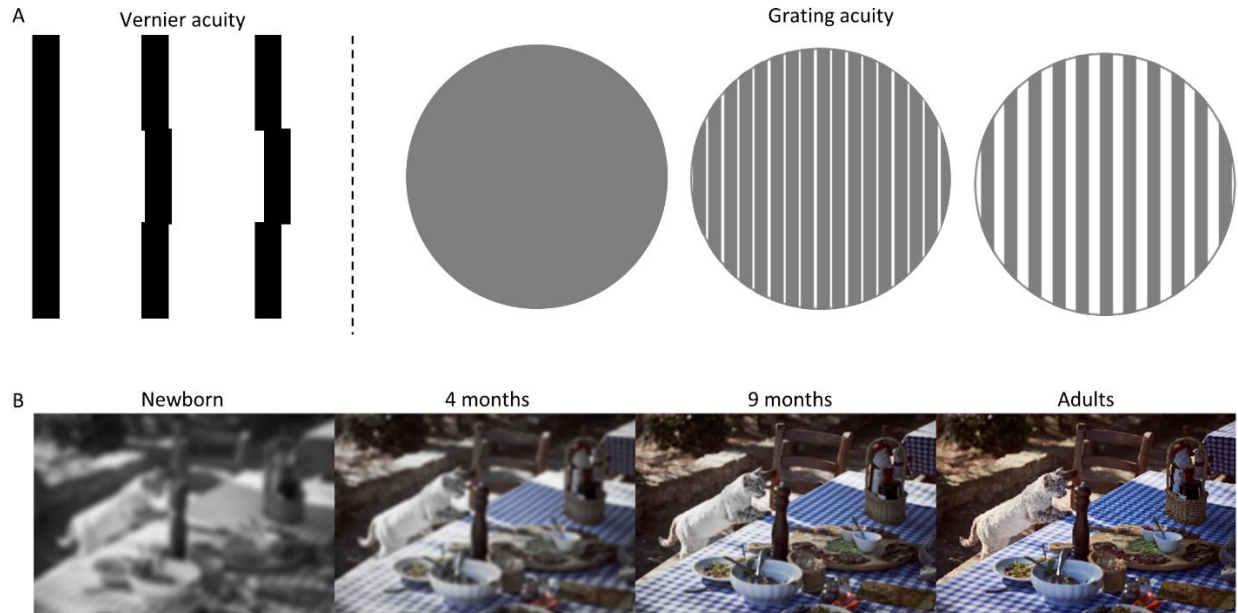


Figure 1: Example of figures use to test infants’ visual acuity and simulation of how infants’ vision. **(A)** Example of Vernier and Grating acuity. For the Vernier acuity, each figure is composed of 3 segments, either perfectly aligned and not discriminable (left), or more and more misaligned and thus easier and easier to discriminate (middle and right). For the Grating acuity, if the gratings are too close, they are not discriminable resulting in the perception of a homogeneous grey patch instead of gratings (left). Increasing the space between the gratings helps discriminate the grated pattern from a homogeneous gray patch (middle and right). **(B)** Simulation of how newborns, 4-month-olds and 9-month-olds see as compare to adults. The acuity is simulated by applying a Gaussian filter, and the saturation of the color was decrease to estimate how infants see according to their threshold of saturation (Knoblauch et al., 2001).

1.3.2. Behavioral evidence of early categorization

Visual categorization in infancy has traditionally been studied using specific paradigms of familiarization or habituation followed by preference-for-novelty (see for review Aslin, 2007; Fantz, 1964; Hunter & Ames, 1988; Rose et al., 1982). In these paradigms, infants are presented with a series of different images belonging to one category (e.g., cats). After a number of familiarization trials, or after they habituated to the category (e.g., look three images in a row for less than 50 % compare to the first three images), they are presented with test trials, consisting in either one image of another category (e.g., dogs) followed by a new instances of the familiarized/habituated category (e.g., a new cat), or by both images presented at the same time next to each other, one being the new category, the other the new instances of the familiar category. If infants’ gaze is attracted by the new category (e.g., looking longer to the dog, but not to the new cat), this is taken as evidence for the infants’ ability to discriminate between the two categories. Another paradigm to study categorization in infancy is the sequential

touching task, but as this task demand a higher motion ability, it is mainly used at the end of the first year onward (Arterberry & Bornstein, 2012; M. H. Bornstein & Arterberry, 2010; Poulin-Dubois et al., 2015).

Using the looking time exploration, studies have shown that already at birth, newborn show a preference for faces and face-like stimuli (two dots above one dot in an eye-mouth configuration) over inverted faces (Farroni et al., 2005; M. H. Johnson et al., 1991; Valenza et al., 1996), for canonical over distorted human bodies (Bhatt et al., 2016; Slaughter et al., 2002; Zieber et al., 2010), as well as for biological motion over non-biological (random) motion (Bardi et al., 2011; Simion et al., 2008). Later on, 4-month-old infants are able to discriminate cats and dogs from birds, cats from dogs, cats from tigers and horses, and zebras from horses (Eimas & Quinn, 1994; see for review Mareschal & Quinn, 2001; Quinn et al., 1993, 2001). However, they were not looking longer at female lions when familiarized to cats (Eimas & Quinn, 1994), nor to a new breed of cat or dog when familiarized to a specific breed (e.g., Siamese or Beagle; Quinn, 2004) suggesting that some categories might be more difficult than others to represent (Mervis & Pani, 1980), and evolved with age. Indeed, 6- to 7-month-old infants did discriminate female lions from cats when familiarized with cats (Eimas & Quinn, 1994), and new breeds from familiarized ones (Quinn, 2004), while younger infants, at 2 months of age, were not even able to discriminate cats from other mammals (Quinn & Johnson, 2000).

Infants can also represent bigger, more global (superordinate) categories, such as mammals. Indeed, 3- to 4-month-old infants familiarized with different exemplars of (terrestrial) mammals (e.g., cat, dog, rabbit, elephant...) habituated to mammals and were thus not attracted by a new exemplar of another mammal (e.g., a deer), but were distinguishing them from non-mammalian animal (e.g., bird and fish), looking longer again when those animals were presented to them. They also distinguished mammals from furniture such as chairs (Behl-Chadha, 1996). In fact, it appears that infants would start to form superordinate categories before basic and subordinate ones, shifting from global-to-basic-level (Pauen, 2002; Quinn & Johnson, 2000). In addition, young infants have different expectation for animate stimuli compare to inanimate stimuli, suggesting a representation of objects by animacy, already early on. Indeed, although later on, preschooler children expect an animal-like agent, able to move by itself, to not be empty, while they have no such expectation for an apparent inanimate stimuli, only moved by a human agent (Kominsky et al., 2021). Even earlier, 3- to 5-month-old infants are not surprised by an animate stimulus moving by itself, or for a self-propelled objects to have goals, while they are not expecting such thing for inanimate objects (Baillargeon et al., 2009; Kominsky et al., 2021; Kuhlmeier et al., 2004; Luo et al., 2009; Luo, 2011; Luo & Baillargeon, 2005; Rakison & Poulin-

Dubois, 2001; Saxe et al., 2005; Spelke, 1994; A. Woodward, 1998). However, they expect inanimate objects to have a cohesion and not randomly split, to move on a continuous path, to be solid and to fall if not supported (Baillargeon et al., 1985, 2008; Kuhlmeier et al., 2004; Needham, 1997; Needham & Baillargeon, 1993; Spelke et al., 1992; Spelke, 1994; Spelke, Kestenbaum, et al., 1995; Spelke, Phillips, et al., 1995). Finally, around 6 to 8 months of age and later, infants start to identify plants as being edible, on the contrary to artifacts, although they also exhibit reluctance to spontaneously touch plants compare to other artificial objects and will look more at adults surrounding them before touching plants (C. Elsner & Wertz, 2019; Wertz & Wynn, 2014a, 2014b).

The behavioral studies thus show evidence for an early categorization of objects, based on visual features (Quinn, 1987; Spencer et al., 1997). However, the number of categories studied was very limited (because of the infant's attentional limitation), and stimuli were repeated a lot, because of the familiarization/habituation phase (Eimas et al., 1994; French et al., 2004). Thus, categories could have been formed online, during the experiment; the different experiments thus tested how easy a category could be formed, more than whether the category was represented by infants. It is unclear whether those categories are spontaneously used by infants, or whether they are only represented when learned, after a long enough exposure to the category. Other studies have thus explored the infants' brain activity, using EEG or fMRI and passive viewing of different images, not relying on an online learning of a category.

1.3.3. Neural evidence of early categorization

Coherent with previous behavioral study, two EEG studies explored 6-month-old infants' ability to learn categories. They both used an event-related potential (ERP) paradigm, exploring how the infants' brain activity changed in function of the stimuli they look at. After being familiarized with cats, 6-month-old infants' brain activity showed a different pattern when looking at exemplar of the familiarized category or of a new category (Quinn et al., 2006). Indeed, the ERP reveal a bigger negative central component (Nc), which has been related to infants' allocation of attention (Reynolds & Richards, 2005), for exemplars of the new category compared to exemplars of the familiarized category. Similar effects were found with a familiarization to bird or fish and test with fish, bird, or car (Grossmann et al., 2009). Those studies evidenced 6-month-old infants' ability to learn to recognize exemplars of a category and discriminate them from a new category.

A similar effect (Nc component) was found in 7- to 8-month-old infants' ERP when looking at animals *vs.* furniture (B. Elsner et al., 2013; Jeschonek et al., 2010), or human *vs.* animal (Marinović et

al., 2014). However, those studies do not familiarized infants to one or the other category, and the Nc component was sometimes bigger for animal exemplars, or, for furniture exemplars. Thus, although this component is not directly related to categorization, it reveals a greater attention from infants to a different category, and evidenced infants' visual categorization between at least animals and furniture, from 7 months of age onward.

In addition, a number of studies have singled out a human face-selective response in the EEG signal of 4- to 6-month-old infants, by tagging the response to human faces appearing in a stream of nonhuman animals and inanimate objects. This response has been found even stronger when infants were in presence of their maternal odor (de Heering & Rossion, 2015; Leleu et al., 2020; Rekow et al., 2020; Yan et al., 2023). The maternal odor even enhances a face-selective response to face-like stimuli, so stimuli that have a similar structure than a face (2 eggs and 1 slice of bacon in a pan fry disposed in a face-like configuration) but are not a face (Rekow et al., 2021). Such human face- or face-like-selective response was also described in human adults using the same paradigm (Quek & Rossion, 2017; Rekow, Baudouin, Durand, et al., 2022; Rossion et al., 2015), emphasizing the evidence of an adult-like categorization for human faces already in 4- to 6-month-old infants. Other EEG studies have also found a specific response for human faces over the right hemisphere in 2- to 5-month-old infants (Adibpour et al., 2018; Di Lorenzo et al., 2020). The ERP component of 3-month-old infants also reveal an inversion effect for faces but not for car, revealing the special case of faces, already in infancy (Halit et al., 2003; Peykarjou & Hoehl, 2013). Even earlier, in the EEG signal of newborns, a face-selective response was found by contrasting upright and inverted face-like configuration (two dots above one dot in an eyes-mouth configuration; Buiatti et al., 2019). Besides evidence for face-selective brain responses, ERPs of 3-month-old infants also differs when viewing bodies (or faces) as compared with distorted counterparts (Gliga & Dehaene-Lambertz, 2005), suggesting an ability to discriminate intact from distorted bodies. However, it appears that not all categories develop at the same time, as other studies have found different emergence of various categories. Indeed, by tagging the response to specific categories such as limb, characters, faces, corridors and cars in infants from 3- to 15-month-olds, the first category to emerge in youngest infants was found to be the human face category, in 4-month-old infants, but other category-selective responses were found only later on, from 6 months of age (Yan et al., 2023). Tagging animals in a stream of furniture, or furniture in a stream of animals revealed categorization between animals and furniture in infants, children and adults, as soon as 4 months of age (Peykarjou et al., 2023). The categorization response increase with age, and in infants (4-, 7- and 11-month-olds), only the categorization of animals in a stream of inanimate objects (4- and

11-month-olds) or the reverse (7-month-olds) was found, while both directions were found in children and adults. Phase-scramble counterparts did not elicit categorization response in infants, suggesting a categorization of images based on more higher-level visual features as compared to what is available in phase-scramble stimuli (e.g., low-level features such as color, contrast, luminance, power spectrum).

Decoding analyses can also be employed on infants' ERP signal, to compare the activity recorded for different instances of one category *vs.* different instances of another category, testing whether the overall signal for one category differ from the overall signal of the other category (Grootswagers et al., 2017). Using this method, studies have found that infants, from 6 months of age, can categorize objects such as toys, bodies, houses and faces at their basic-level, although the categorization in the EEG signal of infants emerge later than the one in adults (Xie et al., 2022). As this method allows to extract pattern of activity in the EEG response, this study also suggest that the infants' basic-level categorization is based on rather mid-level visual cues. Indeed, as for adults' fMRI data, they extracted the representation of objects from the infants' EEG signal via the pattern of correlation, and found that infants' representation of objects is similar to the objects' representation extracted from the mid-level layers of a deep neural network (DNN). DNNs are models of the human adults' visual cortex, and mid-level layers are predictive of the objects' representation in mid-level visual areas, that encode mid-level visual features. Thus, they concluded that infants' representation of objects is based on the mid-level features when representing the basic-level category. Later on, 15-month-old infants also distinguish objects such as cat, dog, bunny and teddy bear from each other, as well as object part such as nose, foot, hand and mouth (Bayet et al., 2020). However, contrary to adults, they seem to not represent the more superordinate categorization (i.e., nonhuman animals *vs.* human body parts). Indeed, in adults, the decoding (i.e., discrimination via a support vector machine (SVM) that is trained to distinguish between two or more patterns) of the EEG signal between superordinate categories, such as hand *vs.* dog, elicit a higher accuracy (i.e., better performance of the SVM) than the decoding within superordinate categories, such as hand *vs.* nose. This is suggesting that body parts and nonhuman animals are eliciting different responses in the adults' brain, leading to bigger differences between the EEG signals, thus an easier decoding, as compare to within nonhuman animal and within body part categories. For 15-month-old infants, the difference of the brain response for a nonhuman animal *vs.* a body part is not larger than for a nonhuman animal *vs.* another nonhuman animal, or for a body part *vs.* another body part, suggesting no representation of categories between nonhuman animal and human body part, only the representation of the object at their basic-level (in this study, it is equivalent to the identification of a stimuli, as only 8 images were used). However, by 15 months of age, the decoding

between categories emerge at similar timing than in adults (i.e., the brain responses start to differ at similar timing), although the peak of the decoding (i.e., the timing at which the brain responses are best discriminate from each other) is still decade as compare to adults' one (Bayet et al., 2020).

As adults' fMRI studies have evidenced specific, selective activity for categories in the ventral stream (Downing et al., 2001, 2006; Epstein & Kanwisher, 1998; Grill-Spector & Weiner, 2014; Kanwisher et al., 1997; McCarthy et al., 1997; Pitcher et al., 2009; Reddy & Kanwisher, 2006), researchers have investigated the organization of part of the infants' ventral stream, the extrastriate cortex, using fMRI in awake infants while looking at different videos of objects, faces, bodies, or scenes. As the adults' extrastriate cortex present hotspots of activation for stimuli such as faces, bodies and scenes, researchers have tried to evidence the same hotspot in infants. They found areas, spatially localized in places very similar to what is typically found in adults, showing some selectivity for faces *vs.* scenes, and scenes *vs.* faces, although the selectivity of those hotspot did not generalize when contrasted with objects instead of scenes or faces (Deen et al., 2017). Later on, those local hotspots were again identified in other 4- to 6-month-old infants, in addition to another hotspot, for bodies. This time, data show, for each hotspot, voxels selective for the corresponding category, so voxels selective for faces *vs.* everything else in the hotspot selective for faces, for bodies *vs.* everything else in the hotspot selective for bodies and for scenes *vs.* everything else in the hotspot selective for scenes (Kosakowski et al., 2022). Those category-selective responses might actually be anticipated, or made possible, by the functional connectivity within face-selective visual areas and within scene-selective visual areas. Indeed, already in neonates, functional connectivity between regions localized in areas corresponding to face-selective adults' areas and to scene-selective adults' areas, across the ventral stream, was evidenced (Kamps et al., 2020). In adults, an area selective for faces is found in the occipital cortex, in addition to the one typically found in more anterior part of the ventral stream, in the ventral occipito-temporal cortex (VOTC). Both these face-selective areas are highly functionally connected one another, meaning that evidences of anatomical and functional connections (Gschwind et al., 2012; Pyles et al., 2013; Zhang et al., 2009), as well as causal activation in adults exist. Indeed, when the area in the occipital cortex is de-activated by a pulse of magnetic stimulation, the more anterior face-selective area is also disrupted (Pitcher et al., 2014). Similar organization was also found between two place(scene)-selective areas within the adults' visual cortex (Gschwind et al., 2012). These functional connections between selective regions were demonstrated in areas localized in similar places in the neonate's brain, suggesting that the local hotspot evidence in older infants (4- to 6-month-old infants) in fMRI might actually be already in place at birth, or are formed via those pre-existent connection

between those areas (Kamps et al., 2020). The similarity between infants' and adults' spatial organization in the ventral stream was also evidenced in animal studies, that measured nonhuman primate inferior temporal activity using fMRI from birth, and found a remarkable spatial organization in the fMRI activity that correspond to what was already evidenced in adult monkeys (Arcaro & Livingstone, 2017), with spatially similar areas also selective for bodies (Arcaro et al., 2019) and faces (Livingstone et al., 2017).

1.3.4. Summary of the early visual categorization

Overall, studies of infants' brain activity and behavior have suggested an early (i.e., found in young infants) categorization of objects, although emerging slower than in adults (ERPs studies) and based on more mid-level cues compare to adults' one, as suggested by similarities with different layers of DNNs (Peykarjou et al., 2023; Xie et al., 2022). Studies especially revealed the categorization of human faces (Adibpour et al., 2018; de Heering & Rossion, 2015; Di Lorenzo et al., 2020; Halit et al., 2003; Kosakowski et al., 2022; Leleu et al., 2020; Peykarjou & Hoehl, 2013; Rekow et al., 2020, 2021; Xie et al., 2022; Yan et al., 2023), and emergence of other categories such as animals *vs.* furniture (B. Elsner et al., 2013; Jeschonek et al., 2010; Peykarjou et al., 2023) from 4 months of age. Other, more basic-level categories seem to emerge later on in infancy (Marinović et al., 2014; Quinn et al., 2006; Xie et al., 2022; Yan et al., 2023). Yet, due to difficulties in infants' testing, mainly caused by the short attentional span infants have (Aslin & Fiser, 2005), even the studies aiming at exploring infants' overall categorization and organization of visual information in the infants' visual system were often limited to quite narrow categories (i.e., faces, bodies or body parts, toys, places, few mammals, corridors, houses, cars), or in the number of instances within category (or both), and usually including faces and bodies or body parts (Bayet et al., 2020; Kosakowski et al., 2022; Xie et al., 2022; Yan et al., 2023; See Peykarjou et al., 2023 for exception), that do not account for the overall organization present in the adults' visual ventral stream, where we found an organization between animate and inanimate stimuli, that crumbles into finer-grained distinction between human and nonhuman faces and bodies and natural and artificial big and small objects (Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; see for review Op de Beeck et al., 2008). Thus, infants' studies have mainly focused on human and natural big (i.e., natural scenes) or some basic-level artificial big or small objects (i.e., corridors, houses, cars or toys) but without collecting the whole categories together. Especially, categories of animal, such as birds, mammals or fish, and natural small objects such as fruits, vegetables or flowers, while being

relevant for humans' everyday life, as being potentially predators and/or foods, have been poorly investigated, for the benefit of mainly human faces.

To better understand the overall development of visual categories and their influence on humans' behavior, understanding their development, the timeline of their emergence in infancy and their influence on infants' visual exploration appears to be fundamental. If studies of infants' behavior have so far informed us about infants ability to form online categories (see for review Mareschal & Quinn, 2001) and their reliable preference for faces or face-like stimuli (Farroni et al., 2005; M. H. Johnson et al., 1991; Valenza et al., 1996), it is not yet clear which category would drive infants' exploration without previous specific exposure, mirroring closer the infants' exploration outside of the lab. On top of this matter, the respective roles of brain maturation and visual experience in the spontaneous representation of visual categories remains to be elucidated; so far, studies have evidenced selective areas, similar to what is find in adults, with infants having limited visual experience, suggesting a role of the brain maturation. However, experience might still act in this organization, as not all categories was yet found in the fMRI activation of the infants' brain, and as infants had still some visual experience when tested. Thus, studies have yet to elucidate whether the brain maturation limits the brain organization by visual categories, or whether visual experience majoritarilly acts on this organization, in order to better apprehend when infants start to see the world as adults do. Moreover, larger categorization such as the big animate-inanimate categorization has still to be characterized in young infants, including more within category variability, as studies focused more on narrower categorization such as mammals or animals *vs.* furniture. Finally, if visual categories are the product of all the available visual features of the object, the specific role and weight of lower-level features in the formation of visual categories, and especially in the big animate-inanimate distinction, remains unclear.

This thesis aims at investigate the emergence in infants' visual exploration of a large number of categories, representative of the ones organizing the adults' ventral stream, and the role of the brain maturation in this emergence. Then, focusing on the big animate-inanimate distinction, I will investigate further the difference between infants' and adults' categorization capacity, using the larger stimuli set so far for this categorization, and investigating the automaticity and limits of the categorical response. Finally, I will study the role of lower-level visual features in the adults' visual categorization.

1.4. Thesis overview

1.4.1. Current research

Investigating the development of visual categories is primordial in understanding the organization of humans' cognition, as the visual categorization is at the interface between *seeing* and *thinking* about an object. I have here narrowed this big topic in 4 smaller studies, all aiming at understanding how the organization found in the human adult's ventral stream comes about. First, I will investigate the emergence of the visual categories that account for the adults' ventral stream organization in the first 2 years of life. Second, I will explore the respective role of brain maturation and visual experience in the previously described emergence of visual categories. Third, I will study the infants' limitation in the animate-inanimate categorization. Finally, I will investigate the resilience of the adults animate-inanimate categorization.

1.4.2. Overview of chapters

Chapter 3 (study 1) investigates how visual categories influence infants' visual exploration, without any prior experience of any images. We explored the emergence of 8 basic-level categories that all encompass the big organization observed in the human adults' ventral stream, using infants' looking time: human and nonhuman animal faces and bodies, and natural and artificial big and small inanimate objects. Because relevant for the organization of the adults' ventral stream, as well as the organization of the primate visual system in general (Kriegeskorte, Mur, Ruff, et al., 2008), those categories can indeed be the earliest to account for infants visual exploration in the first 2 years of life. We presented images by pair of within or between category images to infants and measure their looking time towards each image. Using representational similarity analysis (RSA), we compare infants' looking time to models representing the adults' organization typically find in the ventral stream, as well as directly to the adults' organization in their ventral stream, as measured by fMRI. Overall, this chapter show two milestones in the infants' visual categorization development. First, infants' reliance on non-categorical features (i.e., features that are not eliciting the representation of a category) when exploring objects, although already representing the animate-inanimate categorization. Then, the emergence of more and more visual categories, while older infants relied more and more on category membership when exploring objects, better fitting the adults' organization of the visual ventral stream. Altogether, it suggests that as infants grow older, they represent more and more features when observing objects, allowing them to better represent those objects and form finer-grained categories.

Chapter 4 (study 2) investigates further the reliance on category membership over non-categorical features in infants' visual exploration, and the roles of brain maturation and experience. It replicates the previous study, refining the transitional age at which the reliance on category membership emerge spontaneously in infants' behavior, and investigates the respective roles of brain maturation and experience by generalizing previous results to a new population, preterm infants. Results show that infants start to explore objects based on their category at 8 months of age, while still be guided by the non-categorical features. Preterm infants reveal the importance of the experience in the emergence of category membership in infants' visual exploration.

Chapter 5 (study 3) investigates the limitation of the automatic categorization of animate and inanimate stimuli in infants and adults, using the largest stimuli set so far. We recorded infants' and adults' brain activity via EEG, tagging the categorization of a large pool of animate and inanimate stimuli, that aims at representing the large variability found in the world. We varied the speed of presentation of the stimuli, to explore the limitation of this categorization (i.e., does it breaks down if higher speed of presentation, similarly in infants and adults). Results show that with the increase of the speed of images presentation, the brain response revealing the categorization decrease, at each stage of development. Throughout life, from infancy to adulthood, this response is captured with increasing speed of presentation: we were able to capture the response with a speed of presentation 3 times faster at 9 months of age compare to 4 months, and even 7 times faster in adults compare to 4-month-old infants. Overall, this study revealed the automaticity of the animate-inanimate categorization throughout development, which is subject to different time limitations throughout the development.

Chapter 6 (study 4) investigates the role of low-level visual features in the big animate-inanimate categorization in adults. We again recorded adults' brain activity via EEG, tagging the categorization of animate and inanimate stimuli, this time varying the stimuli. We impoverished the stimuli so that only mid- to low-level visual features remained. Results show that with all stimuli, even with the most impoverished ones containing only few low-level visual features, the adults' brain is still able to categorized unrecognizable animate and inanimate stimuli. However, the response obtained from very impoverished stimuli show that the low-level features does not account for the overall response obtained with the intact stimuli. Overall, this study shows that if the animate-inanimate categorization is resilient to the loss of visual information, higher-level visual features play a major role in this categorization.

2. General methods

2.1. Overview of the stimuli

2.1.1. Stimuli used in Chapter 3 and 4

The visual stimuli used in chapters 3 and 4 (studies 1 and 2) aimed at increasing the number of visual categories studied at once in infancy, while fitting their attentional limitations. Stimuli encompassed 8 visual categories, aiming at representing the overall organization of the adults' ventral stream. In total, we used 72 color photographs of isolated real-world objects. Most of them were selected from publicly available sets (Kiani et al., 2007), and some were selected from the internet. Objects were 9 instances of human faces, 9 instances of human bodies, 9 nonhuman animal faces (e.g., cat face), 9 nonhuman animal bodies, 9 natural big objects and scenes (landscapes including only natural objects such as mountain and trees), 9 natural small objects (fruits, vegetables, flowers and leaves), 9 artificial big objects (e.g., house, chair) and 9 artificial small objects (e.g., hat, key). The human bodies were all view from the back with no face visible but with the head, to avoid having a categorization of human bodies due to the faces. However, nonhuman animal bodies depicted the face of the animals, although from profile view.

2.1.2. Stimuli used in Chapter 5 and 6

The visual stimuli used in chapters 5 and 6 (studies 3 and 4) aimed at representing as much as possible the huge variability found in the real-world in the animate and inanimate categories. In total, we used 640 color photographs of isolated real-world objects superimposed on a gray background, taken from the internet. Animals ($n = 320$) encompassed mammals, fish, birds, amphibians (frogs) and reptiles (turtles). The mammal category included all type of mammals, except humans, to avoid a categorization of animals only due to the presence of humans. As very similar to humans, primates (e.g., gorilla) were only a few. Categories such as dog and cat, that are very common in our everyday life, were, as much as possible, underrepresented – i.e., there are some instances of cats and dogs, but they were not the majority of mammals. Inanimate objects encompassed big and small natural and artificial objects. This stimuli set was further impoverished in numerous way in Chapter 6 (study 4). The stimuli set had been transform in grayscale, matching the luminance between the animate and the inanimate stimuli, and removing all color information. It had been transformed into texform stimuli (Deza et al., 2019; Long & Konkle, 2017), removing higher-level visual features. The texform transformation preserve the mid-level visual features of the original images, but render the image not

recognizable at the basic-level. Finally, the original set of images and the grayscale version were phase-scrambled, rendering the images not recognizable either, and preserving only low-level visual features such as the power of the original image spectrum or the color (if preserved in the original images). Those transformations aimed at studying the implication of different features in the animate-inanimate big categorization.

2.2. The looking time in infancy

2.2.1. The looking time as a measure of dissimilarity

In Chapters 3 and 4 (studies 1 and 2), the difference of looking time between two images were considered as a measure of the dissimilarity between infants' representation for each image. The infants' looking time is one of the main measure use to study infants' cognition (Aslin, 2007). We used the difference of looking times, assuming that looking times for 2 objects never seen before would be more similar, the closest their visual representation is. Indeed, such an effect is suggested by previous studies investigating how language influence the infants' categorization of objects (Balaban & Waxman, 1996; Ferry et al., 2010; LaTourrette & Waxman, 2020, 2022; Waxman & Markow, 1995). Those studies, like all studies using the habituation and novelty-for-preference paradigm, interpret a difference in looking times between 2 images as representing the infants' ability to distinguish the two images. Often, it was used to test infants' representation of categories, and thus, the differential looking time was not only reflecting the infants' ability to distinguish between the two images presented, but as representing the infants' ability to represent two images as part of different categories. Thus, we hypothesize that, when facing two images never previously seen, infants will have a representation of each of them, and look at them accordingly. If their representation of the two images is similar, they would then look at the two images the same way, and thus, about the same amount of time. On the contrary, if they represent the images as being different (more dissimilar), they would react differently, thus, look at one of the images more than the other. According to this hypothesis, the differential looking time was considered as a measure of the dissimilarity of infants' representation of each image in a pair.

In the typical habituation and novelty-for-preference paradigm, infants' looking times were used to measured differences between a familiar stimuli or a familiar category, *versus* a novel stimuli or a novel category. The familiarity with a stimuli or a category is then created online, during the experimental session, in the lab. Instead, in Chapters 3 and 4 (studies 1 and 2), we used looking times to measure the (dis)similarity perceived between two objects, without any previous exposition to one

stimuli or its category. This was a methodological challenge, but also an innovation in the exploration of infants' looking time to study infants' cognition. We were able to represent infants own spontaneous representation for each object in relation to the others, exploring the organization of categories in their behavior.

2.2.2. The Representational Similarity Analyses

The Representational Similarity Analyses (RSA) is an analysis allowing to compare together data acquired by different ways, such as fMRI and extracellular recording, coming from different measurement (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). It is also allowing comparison between different species, such as between human and another nonhuman primate such as the macaque (Kriegeskorte, Mur, Ruff, et al., 2008). The logic of RSA is to first work out a Representational Dissimilarity Matrix (RDM) for each measurement, that represent the dissimilarity between all the stimuli of the study, and for each measures. Those matrices are then representing, according to each population and/or measure, the dissimilarity between all stimuli. Then, the comparison between matrices allows to determine whether the representation, captured in each population and/or via different measures, are organized similarly, or not. It also allows to test specific hypothesis, by putting together theoretical model representing, according to the challenged hypothesis, the dissimilarity across stimuli.

We thus used the RSA to compare infants' differential looking time with the organization of visual stimuli as measured via fMRI in the adults' visual ventral stream. RDM were computed, for infants' behavior, with the differential looking time, and for adults' representation, computing the correlation between the different pattern of activity obtained for each stimuli, in 3 different region-of-interest (ROI) in the ventral stream, and all along a vector-of-ROI ranging all along the postero-anterior axis of the ventral stream. The adults' ventral stream activity was measured by fMRI, in the form of a blood-oxygenation-level dependent (BOLD) signal, consisting at mapping the level of oxygenation in the brain. Any variation of the BOLD signal is interpreted as reflecting a variation of the neuronal oxygen consumption, thus as a variation of the neural activity. The BOLD signal is recorded at multiple voxels in the brain. A voxel is a three-dimensional rectangular cuboid of a few millimeters. Patterns of activity correspond to the activation (the BOLD signal) of all the voxels in a region for a stimulus compare to a baseline. A region being decomposed into many voxels, the pattern of activity is as many activation values. It corresponds to the way a region gets activated by a stimulus. By computing a correlation between two pattern of activity, each measured for one stimulus, we obtained a measure

of the similarity between the two recorded activities. However, as RDM represent the dissimilarity between stimuli, we simply subtracted the coefficient of correlation (the measure of similarity), a value between 0 and 1, to 1, and obtained a measure of the dissimilarity. Computing those RDM allowed us to compare infants' organization of visual stimuli in their behavior with the organization of the same stimuli in the adults' ventral stream.

The RSA also allows to test specific hypothesis, by the mean of theoretical models. Those models consist of mapping the different dissimilarity between the stimuli, according to hypothesis. For example, if we want to test the specific hypothesis that the stimuli will be represented as either animate or inanimate, it suggests that all comparison within the animate and within the inanimate categories will be less dissimilar than the between animate-inanimate comparisons. To model such a hypothesis, we thus will create a RDM, consisting of 0 (non/low dissimilarity) for each comparison within stimuli of the same category (e.g., all animate stimuli comparisons), and of 1 (very high dissimilarity) for each comparison between 2 categories (e.g., all animate *vs.* inanimate comparisons).

2.3. Tagging a category-selective response in the EEG signal

2.3.1. Basics of electroencephalography (EEG)

Neuronal activity in the brain produces electrical fields that are recordable, especially in human, directly from the scalp. Indeed, neurons function together and communicates via electrochemical signals, that do not only go from one neuron to the other, but is also spreading in the neuron environment, the tissue surrounding the neurons. Those electrical fields formed oscillation, known as neural oscillations. Measuring the neural oscillations as a function of an event allow to study the changes in the brain activity according to the event. The EEG is the measurement of the neural activity via electrodes placed on the scalp (see for review M. X. Cohen, 2017). The EEG recording takes the form of a difference in voltage between two different electrodes placed at different places on the scalp: an electrode of reference, typically placed on the vertex, is usually used to compute the voltage at all other electrodes on the scalp. The voltage measured at an electrode comes from billions of neurons that “fired” (gets activated) simultaneously, producing an electrical signal that spreads in surrounding tissues, and is thus reflecting a neural populations activity instead of one neuron activity. Small-scale and asynchronous activity is thus difficult or impossible to measure in EEG, and the localization of the EEG signal is not better than at the centimeter order (Olejniczak, 2006). However, the temporal scale of the EEG signal is near the actual firing of neurons, as the electrical signal travels very fast in the

tissue to the electrode. Thus, the EEG signal allows to study rapid changes in the brain activity according to specific events, such as viewing specific images from different categories.

2.3.2. The frequency-tagging paradigm

Repetitive stimulation of the human brain, by means of visual or auditory stimulations for instance, at a constant frequency (e.g., 6 stimulations or cycles per second, or 6 Hz) leads to a synchronization of the brain activation (firing of neural population) with this external stimulation (Regan, 1966, 2009). This can be measured via EEG, and produces an electrical response oscillating at the same stimulation frequency (e.g., 6 Hz). When the stimulation is visual, it elicits a response, called steady-state visual-evoked potential (SSVEP) over visual areas (Rager & Singer, 1998; Rossion & Boremanse, 2011). Such a response can be visualized and quantified in the frequency-domain. The EEG signal is a signal recorded over time, that can be decomposed into frequencies and measured in the frequency-domain, by the means of a Fast Fourier Transform (FFT) algorithm. Typically, a high amplitude at the stimulation frequency (named fundamental frequency) and its harmonics (multiple of the fundamental) is measured after a visual regular stimulation.

In the frequency-tagging paradigm, the principle is to measure SSVEP for different stimuli, presented at specific frequencies. For instance, it was mainly used to study the human face categorization. Stream of images randomly selected from different (non-human face) categories (e.g., trees, flowers, nonhuman animal faces, bodies...) are presented at a specific frequency, typically 6 Hz. In this stream of images, every 5 images, a human face is presented, eliciting a face-selective presentation at a frequency of $6/5 = 1.2$ Hz. Not only one human face is presented, but different instances of human faces. Thus, if a synchronization of the brain activity happens to this face-selective frequency, it is interpreted as a categorical response: the brain gets activated similarly between all human faces instances, and differently than for all other images (see for review Rossion et al., 2020). The presentation of the selective category (i.e., in the previous example, human faces) can be referred to as the oddball stimulation, the deviant stimulation or the category-selective stimulation, and is the tagging of a category in the human brain. It has been previously used to study human faces categorization and other categorization in human adults (see for review Rossion et al., 2020) as well as in infants (see for review Peykarjou, 2022).

3. The visual object categorization in infancy (study 1)

This chapter reports and discuss the study published in Spriet et al. (2022). The introduction of this paper is a summary of this thesis introduction, and tackled the very first question: which categories account for how infants explore their environment, influencing human behavior from the very beginning? The supplementary materials of this article can be find in Chapter 9, after the bibliography.

3.1. Introduction

Objects are the units of attention and perception; categories are the units of thought. We see objects (e.g., a rounded spongy red and white-dotted shape on an elongated support), but we think about objects primarily in terms of categories (e.g., the mushroom *Amanita muscaria*). By recognizing an object as member of a category, we understand what that object is and retrieve its visible (e.g., it is red with white spots) as well as its invisible properties (e.g., it is hallucinogenic). Categorization is thus the basis of inference and decision, although not all inferences and decisions require categorization.

Objects can be categorized according to a virtually infinite number of perceptual and nonperceptual dimensions (Anderson, 1991; Rosch, 1978). Insight on the most basic and general dimensions for object categorization in humans has been gained by studying how information is organized in the vast brain territory for visual object representation, which forms the occipitotemporal visual ventral stream.

Here, categories emerge from the topography of responses to visual objects, resolving into a large-scale organization that distinguishes between animate and inanimate objects, and crumbles in finer-grained distinctions between human vs. non-human animals, small vs. big (in terms of real-world size) (Konkle & Caramazza, 2013; Konkle & Oliva, 2012b), and natural vs. artificial objects (M. A. Cohen et al., 2017; Haxby et al., 2011; Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Long et al., 2018; Op de Beeck et al., 2008, 2019; Papeo et al., 2017). Underneath this organization lies a mosaic of local hot spots of strong selectivity for stimuli, such as faces, bodies, and scenes (Downing et al., 2001; Epstein & Kanwisher, 1998; Kanwisher et al., 1997; McCarthy et al., 1997). Because of its organization and role in object recognition, the visual ventral stream is regarded as the interface between perception and cognition, forming the backbone for semantic categorization and representation of object and action knowledge in the rest of the brain (Mahon & Caramazza, 2009).

Besides the topography, categorical distinctions in the visual cortex also emerge from dissimilarity between distributed patterns of neural activity evoked by individual objects (Bobadilla-Suarez et al., 2020; Davis & Poldrack, 2013; Kriegeskorte, Mur, Ruff, et al., 2008). Thus, in visual areas, activity patterns recorded with functional MRI (fMRI) are more similar (i.e., less discriminable) for two animate objects (e.g., parrot and camel) than between an animate and an inanimate object (e.g., parrot and car). Visual object categories represented in the visual cortex prove behaviorally relevant, predicting the way in which individuals parse the visual world. For example, in a visual search for a target-object among a set of distractors, people are faster to discriminate and find a target among objects of a different visual category (e.g., a cat among artificial objects) than among objects of the same visual category: search times increase as neural similarity between target and distractors increases (M. A. Cohen et al., 2017).

The organization of the human visual cortex by object categories appears to be a hallmark in the evolution of the primate brain: it is replicated in the visual cortex of monkeys (Kriegeskorte, Mur, Ruff, et al., 2008; Sigala & Logothetis, 2002) and is resistant to variations of individual visual experience (Mahon et al., 2009; Mattioni et al., 2020; Striem-Amit & Amedi, 2014; van den Hurk et al., 2017). A similar organization across species and conspecifics with different environment and life-long visual experience suggests a neural code optimized by evolution. This line of thinking encourages the hypothesis that object representation in the visual cortex reflects biological constraints and dispositions (Caramazza & Shelton, 1998); as such, it would emerge early in life or even be present at birth.

There is initial evidence for signatures, or precursors, of neural specialization to object categories (faces, bodies, animals, and scenes) in the visual cortex of newborns or young infants, based on electroencephalography (Adibpour et al., 2018; Buiatti et al., 2019; de Heering & Rossion, 2015; Gliga & Dehaene-Lambertz, 2007; Peykarjou et al., 2014, 2017) or fMRI (Deen et al., 2017; Kamps et al., 2020). Behavioral counterparts of those neural effects include early preference for faces or face-like stimuli over inverted faces (Farroni et al., 2005; M. H. Johnson et al., 1991; Valenza et al., 1996), for biological over nonbiological motion (Bardi et al., 2011; Simion et al., 2008), and for canonical over distorted bodies (Bhatt et al., 2016; Slaughter et al., 2002; Zieber et al., 2010).

While preference implies discrimination between two objects, visual categorization entails the ability to use the visual properties of a category (e.g., shape) to identify its members and keep them separate from other categories. By 4 mo, infants are already able to do so: exposed to various exemplars

of a category (e.g., cats), they exhibit a novelty effect, looking longer at an object of a new category than at a novel object of the same category (Pauen & Peykarjou, 2021; Quinn et al., 2001; Quinn & Eimas, 1996).

But when do infants begin to see the visual world as adults do? Here, we investigate whether the categorical dimensions that drive the large-scale organization of the human visual cortex could account for the spontaneous emergence and development of real-word object categories in infancy. In particular, under the hypothesis that the structuring of visual object information toward an adult-like organization begins at birth (Buiatti et al., 2019; Deen et al., 2017; Kamps et al., 2020), we asked when such organization becomes functional so as to account for how infants explore the visual world.

We examined the development of visual object categorization in infancy, considering, in one experimental design, objects that have highlighted categorical representations in the visual cortex of human adults (and monkeys): animate vs. inanimate, human vs. nonhuman (animate), faces vs. bodies, natural vs. artificial inanimate, and real-world big vs. small (inanimate) (Kriegeskorte, Mur, Ruff, et al., 2008). Each of the above distinctions defines a categorization model, whereby a given (behavioral or physiological) correlate of object perception would be more similar for two objects of the same category than for two objects of different categories.

Using eye tracking, we recorded the most reliable and informative measure of infants' cognition thus far: the looking behavior (Aslin, 2007; Hochmann, 2013). Infants of 4, 10, and 19 mo viewed two objects at a time on a screen, while we measured the looking time toward either object. We took the looking time difference between two stimuli as a measure of dissimilarity, under the assumption that looking times for two objects seen for the first time would be more similar, the closer their visual representation is (see also ref. LaTourrette & Waxman, 2020). Since two stimuli of the same visual category are, normally, more similar than two stimuli from different categories, we expected the variations in differential looking times (DLTs) to reflect variations in representational similarity, uncovering categorical distinctions.

In classic categorization studies, infants' looking times are used to capture differences in novelty/familiarity, created ad hoc within the experimental session [e.g., through the presentation of multiple exemplars of a category during familiarization (Pauen, 2000; Pauen & Peykarjou, 2021; Quinn et al., 2001; Quinn & Eimas, 1996, 1998)]. Thus, a methodological challenge (and innovation) of the current work was to use looking times to capture differences in the perceived (dis)similarity between two objects, in the absence of any controlled unbalance in the exposure to a given category (at least

within the experimental session). As a result, this approach defined a model where each object was represented in relation to the others (i.e., how similar/dissimilar it was from exemplars of the same and different categories). A model based on a relative measurement can be quantitatively compared with any model based on another relative measurement, whatever the source of the measurements (e.g., reaction times, neural activity) (Kriegeskorte, Mur, & Bandettini, 2008). We compared the model of visual object representation emerging from the infants' looking behavior, with synthetic (i.e., hypothesis-driven) and data-driven (i.e., fMRI-based) models reflecting visual object representation in the mature visual cortex. This approach had previously allowed connecting data from brain-activity recording, behavioral measurements in adults, and computational modeling (Kriegeskorte, Mur, & Bandettini, 2008). Here, by studying the relationship between the infants' looking behavior and the organization of visual object information in the adults' brain, we connected another branch, which is another step toward a unified theory of the origin and development of functional organization in the human brain.

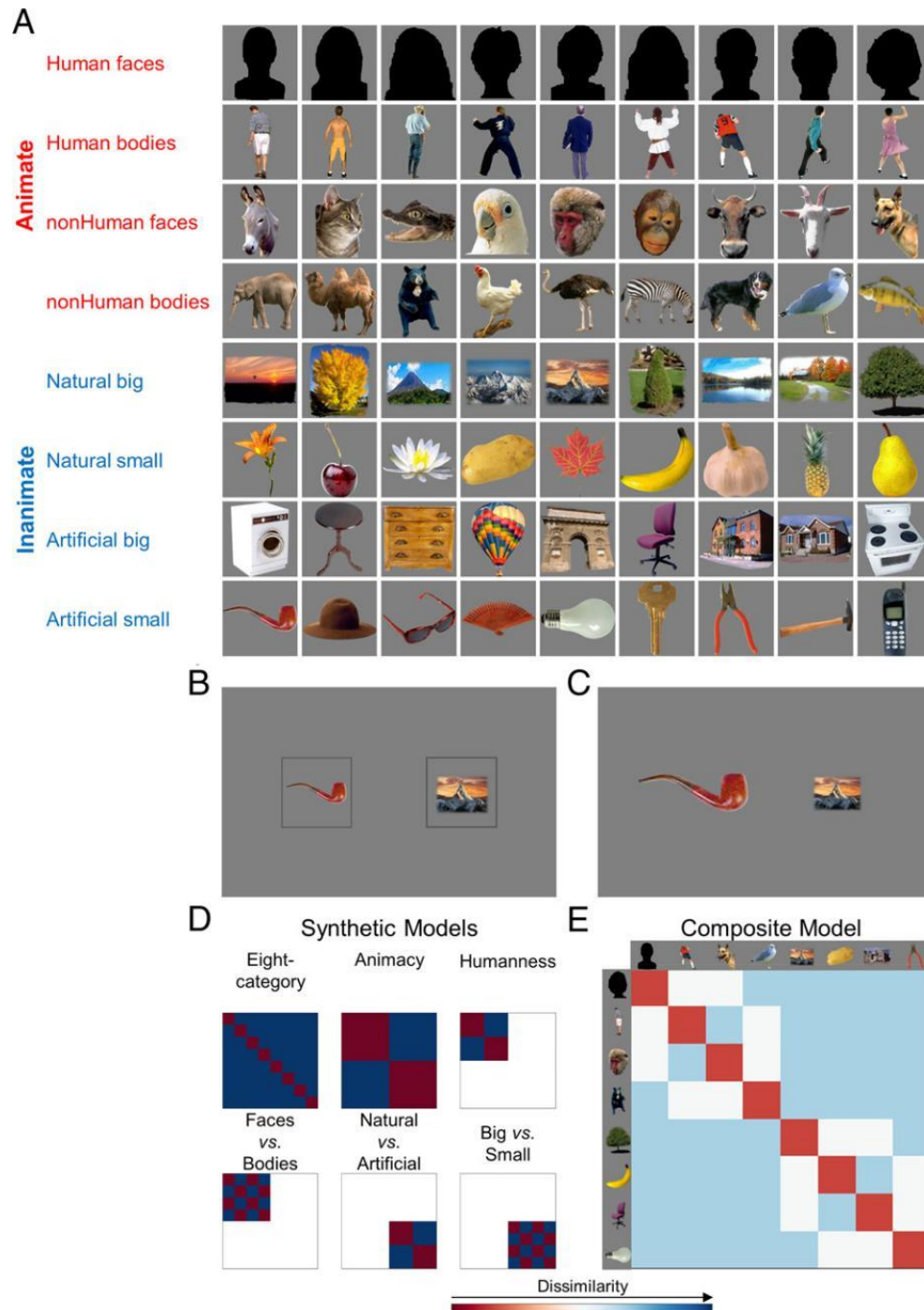


Fig. 1. Stimuli, trials, and hypothesis-based models of categorization considered in the design of Exps. 1 and 2. **(A)** Stimuli were 72 images depicting 9 objects from each of 8 different categories. Silhouettes instead of the actual colorful female human faces used in the experiments, are shown for illustration. **(B)** In each trial of Exp. 1, two images were presented within two gray frames of identical size, on the right and on the left, equally distant from the center of the screen. **(C)** In each trial of Exp. 2, the image frame was removed and the image size was modified so that each object had the same number of pixels. **(D)** Hypothesis-driven (synthetic) models reflecting the categorical object representations tested in the current design. **(E)** The composite model reflecting the mean of the six synthetic models.

3.2. Results

3.2.1. Experiment 1

Three groups of infants of 4 ($n = 24$), 10 ($n = 24$), and 19 mo ($n = 25$) saw 36 pairs of images, each featuring an object from one of eight categories: human faces, human bodies, nonhuman faces, nonhuman bodies, natural-big and natural-small objects, artificial-big and artificial-small objects (hereafter, “big” and “small” refer to real-world size) (Fig. 1A). The set of images used here had previously been used to study object representation in the visual cortex (Kiani et al., 2007; Kriegeskorte, Mur, & Bandettini, 2008). They depicted naturalistic views of real-world objects without meeting any specific requirement or manipulation. Objects thus reproduced the natural combination of visual features typical of their category, so that we could expect the variations in the objects’ appearance to reflect the natural variations within and between categories. All subjects saw all possible combinations (Fig. 1B) of between-category and within-category pairs. For each infant, for each pair, we measured the absolute difference in looking times between the left and right images (DLT) (*Materials and Methods*). DLTs were used to build a representational dissimilarity matrix (RDM), in which cells off the diagonal represented between-category comparisons and cells on the diagonal represented within-category comparisons (Fig. 2). Since different infants saw different exemplars for each category, group-averaged RDMs represented relationships (i.e., dissimilarities) between categories, rather than between individual objects.

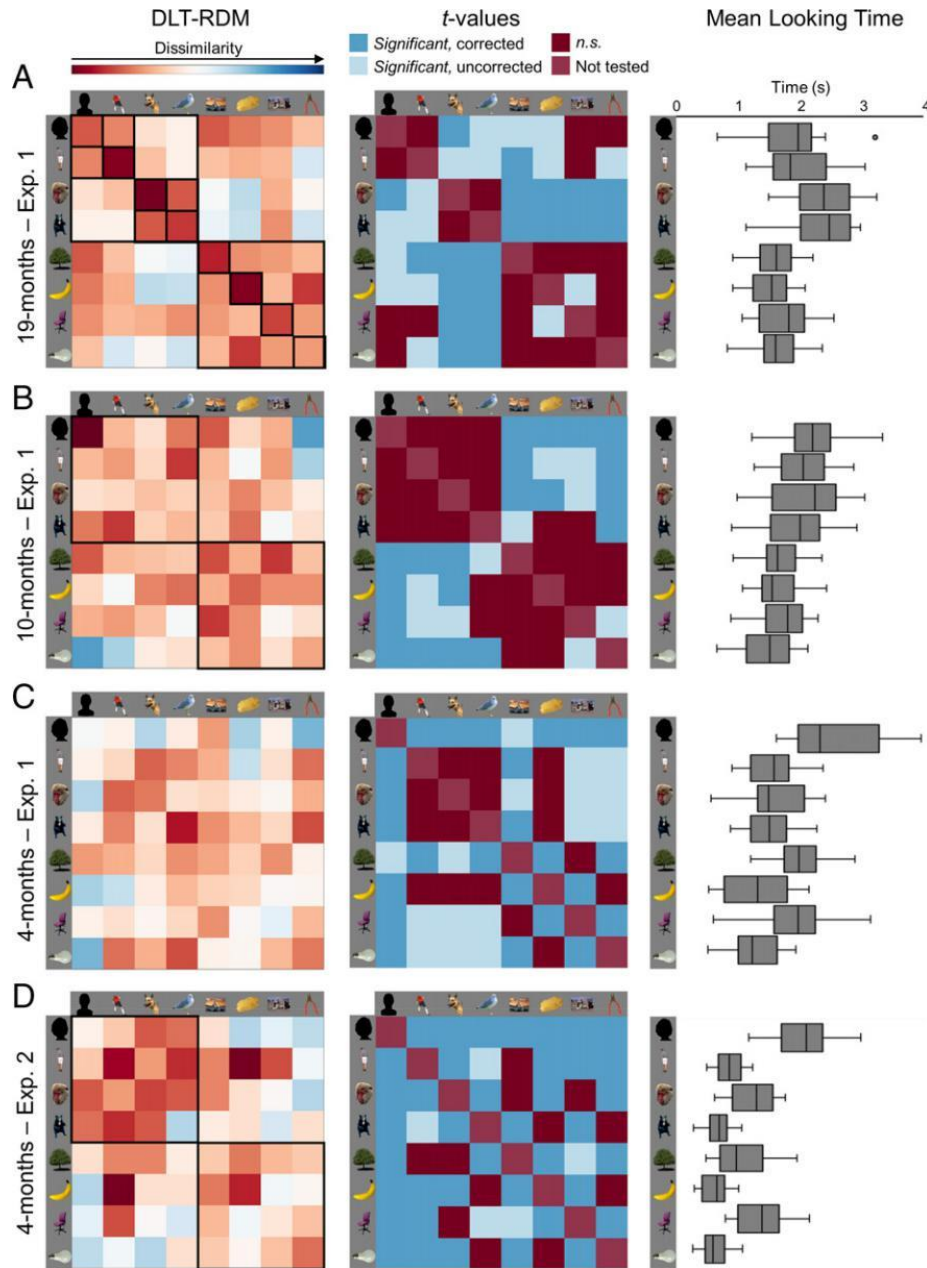


Fig. 2. Results of representational similarity analysis and of the pairwise comparisons of MLT between- and within-categories for each age group in Exps. 1 and 2. **(Left)** Mean RDM reflecting dissimilarities between- and within-categories in terms of DLTs. Black squares in the RDMs highlight categorization by animacy, humanness, and by the eight categories in 19-mo-olds **(A)**, categorization by animacy in 10-mo-olds **(B)**, and in 4-mo-olds of Exp. 2 **(D)**. **(Center)** Matrix of t-values for each pairwise comparison between MLTs of the individual categories for 19- **(A)**, 10- **(B)**, and 4-mo-olds **(C)** in Exp. 1 and 4-mo-olds in Exp. 2 **(D)**. Squares in dark blue denote significant effects; squares in lighter blue denote effects that did not survive the multiple comparison correction (trends); red squares denote nonsignificant (n.s.) or nontested comparisons. **(Right)** Distribution of MLTs in 19- **(A)**, 10- **(B)**, and 4-mo-olds **(C)** of Exp. 1 and of 4-mo-olds of Exp. 2 **(D)**. Box-plots represent the minimum, the first quartile, the median, the third quartile and the maximum of the population distribution; outliers are denoted by dots (one in the 19-mo-olds group).

Reference models of visual object categorization. Using representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008), we computed the relationship between RDMs based on infants' DLT (DLT-RDMs) and models (i.e., RDMs) of visual object categorization in adults, defined with two independent approaches. The first approach defined a set of categorization models based on fMRI responses evoked in human adults, when viewing the same objects presented to infants. In the fMRI-based RDMs, pairwise between- and within-category dissimilarities reflected correlations between neural activity patterns (*Materials and Methods*). Three RDMs were computed from activations in three broad regions-of-interest (ROIs) (Fig. 3 A and B) of the visual cortex (early visual cortex, EVC; ventral occipitotemporal cortex, VOTC; and lateral occipitotemporal cortex, LOTC), and at each location along the antero-posterior axis of the visual ventral stream (i.e., vector-of-ROIs analysis). The second approach defined six synthetic categorization models (RDMs) that may apply to the current stimulus set: animate–inanimate (animacy model), human–nonhuman animates (humanness model), faces–bodies, natural–artificial inanimates, big–small inanimates, and eight-category model, where each category was defined as a category of its own, distinct from the other seven (Fig. 1D). In each cell of an RDM, the values 0 or 1 indicated dissimilarities within-category (lowest dissimilarity) and between-category (highest dissimilarity), respectively. As a model of visual object categorization in adults, a composite-RDM was obtained by averaging the above six models (Fig. 1E).

In addition to, or instead of, categorical information, infants' look might be guided by physical properties of the stimuli, such as size of the image on the retina (Brannon et al., 2006; Libertus et al., 2013), elongation (C. Newman et al., 2001), compactness (Huang, 2020), and color (C. Newman et al., 2001), among others. To assess systematic relations between looking times and visual features of the images, irrespective of the category, we computed RDMs representing differences in size, elongation, and compactness, relying on signed values to appreciate the looking-time difference between two objects but also which one was looked at the longer (the larger one; the more/less elongated one; the more/ less compact one). A fourth RDM was computed to represent differences in the image color (*SI Appendix*, Fig. S1). Note that other image properties may affect infants' behavior; here, size, elongation, compactness, and color should be taken as proxies of unspecific physical properties of our stimuli that are not necessarily distinctive of a category, but could affect the looking behavior.

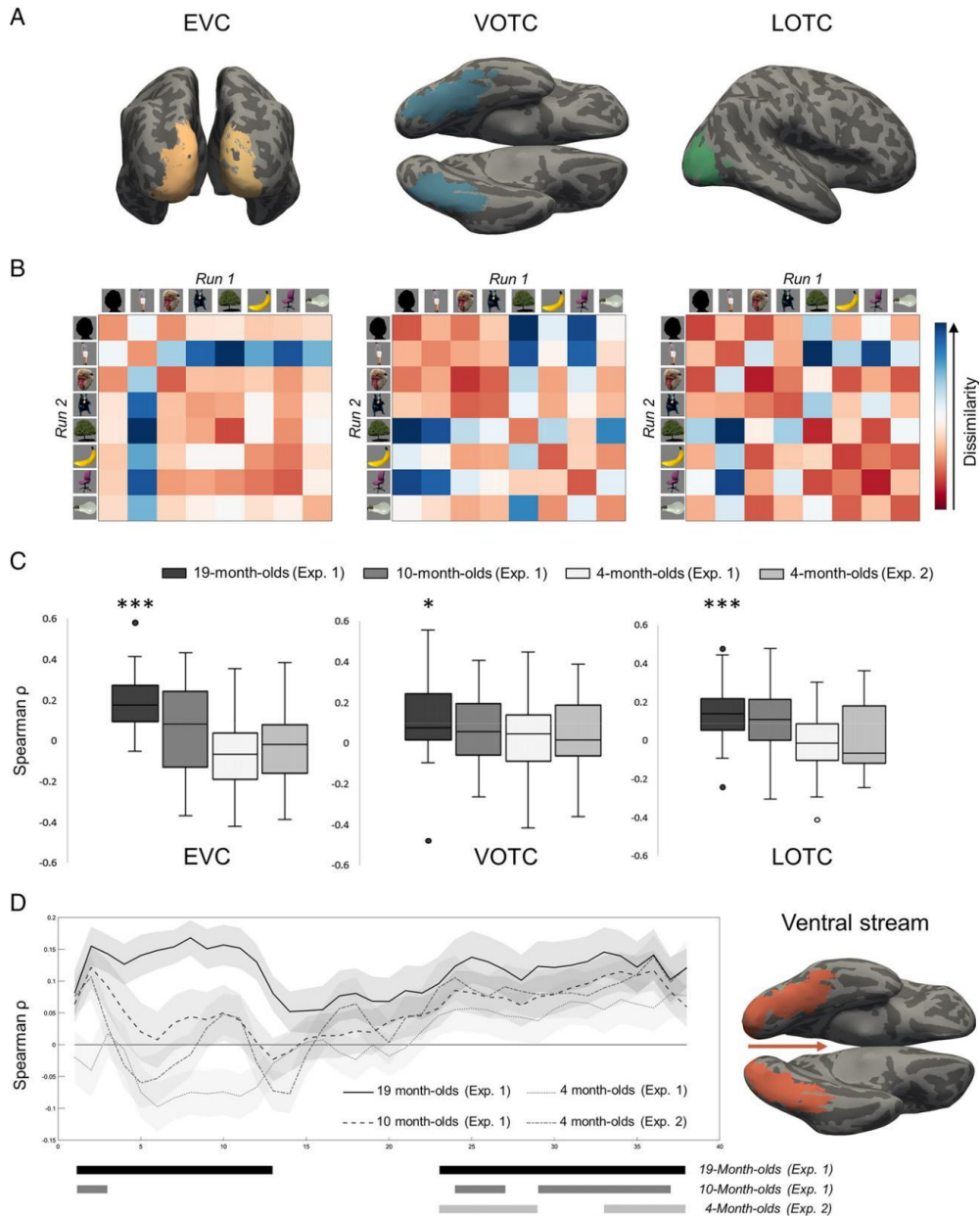


Fig. 3. Relationship between infants' looking behavior and visual object representation in the adults' visual cortex. **(A)** ROIs in the adults' brain: EVC, VOTC, and LOTC. **(B)** Mean RDMs reflecting relationships (i.e., dissimilarities) between object categories in terms of dissimilarities in the neural activity patterns evoked by viewing objects in the EVC, VOTC, and LOTC of adults (fMRI-based RDMs). **(C)** Results of the representational similarity analysis between the mean fMRI-based RDM in each ROI and the DLT-RDM of each infant in each age group of Exps. 1 and 2. Box-plots represent the minimum, the first quartile, the median, the third quartile, and the maximum of the population distribution as well as outliers (dots); $*P < 0.017$; $***P < 0.0003$. **(D)** Results of the representational similarity analysis between the infants' DLT-RDMs and the fMRI-based RDM derived from each partition along the ventral visual stream. Solid bars represent clusters with significant correlation (above 0) for each age group of Exps. 1 and 2.

Table 1. Results of the representational similarity analysis reflecting the relationship of the infants' DLT-RDMs with the fMRI-based RDMs, the RDM for the synthetic composite model of categorization, and the RDMs based on size, elongation, compactness, and color differences.

Exp.	Age (mo)	Model	Mean ρ (SD)	CI (minimum to maximum)	t (df)	P	Cohen's d
1	19	EVC	0.182 (0.123)	0.123 to 0.273	6.795 (24)	<0.0001	1.474
		VOTC	0.110 (0.200)	0.017 to 0.227	2.985 (24)	0.006	0.552
		LOTc	0.133 (0.147)	0.065 to 0.226	4.615 (24)	<0.001	0.902
		CM	0.248 (0.195)	0.168 to 0.328	6.371 (24)	<0.0001	1.272
		Size	0.034 (0.194)	0.139 to 0.071	0.874 (24)	n.s.	0.175
		Elongation	0.010 (0.151)	0.092 to 0.071	0.333 (24)	n.s.	0.066
		Compactness	0.074 (0.166)	0.163 to 0.016	2.218 (24)	0.036	0.444
		Color	0.009 (0.186)	0.092 to 0.109	0.234 (24)	n.s.	0.048
1	10	EVC	0.053 (0.252)	0.079 to 0.185	1.029 (23)	0.314	0.210
		VOTC	0.077 (0.179)	0.017 to 0.171	2.112 (23)	0.046	0.431
		LOTc	0.086 (0.200)	0.019 to 0.191	2.102 (23)	0.047	0.429
		CM	0.125 (0.205)	0.038 to 0.211	2.984 (23)	0.007	0.610
		Size	0.068 (0.208)	0.048 to 0.183	1.592 (23)	0.125	0.327
		Elongation	0.041 (0.197)	0.150 to 0.068	1.013 (23)	n.s.	0.208
		Compactness	0.054 (0.190)	0.052 to 0.159	1.379 (23)	0.181	0.281
		Color	0.044 (0.253)	0.096 to 0.183	0.843 (23)	n.s.	0.174
1	4	EVC	0.067 (0.201)	0.172 to 0.039	1.629 (23)	0.117	0.333
		VOTC	0.033 (0.174)	0.058 to 0.124	0.929 (23)	n.s.	0.190
		LOTc	0.025 (0.184)	0.122 to 0.071	0.673 (23)	n.s.	0.137
		CM	0.049 (0.182)	0.028 to 0.126	1.320 (23)	0.200	0.269
		Size	0.315 (0.173)	0.220 to 0.411	8.950 (23)	<0.0001	1.821
		Elongation	20.321 (0.187)	20.425 to 20.218	28.401 (23)	<0.0001	1.717
		Compactness	0.208 (0.155)	0.123 to 0.294	6.597 (23)	<0.0001	1.347
		Color	0.035 (0.204)	0.078 to 0.148	0.836 (23)	n.s.	0.172
2	4 m	EVC	0.026 (0.199)	0.131 to 0.079	0.642 (23)	n.s.	0.131
		VOTC	0.051 (0.191)	0.050 to 0.151	1.295 (23)	0.208	0.264
		LOTc	0.017 (0.184)	0.080 to 0.114	0.453 (23)	n.s.	0.092
		CM	0.102 (0.205)	0.015 to 0.189	2.433 (23)	0.023	0.498
		Elongation	20.150 (0.187)	20.253 to 20.046	23.924 (23)	<0.001	0.802
		Compactness	0.269 (0.221)	0.146 to 0.391	5.949 (23)	<0.001	1.214
		Color	0.068 (0.182)	0.033 to 0.169	1.837 (23)	0.079	0.374

CM, composite model; mean ρ are the Fisher-transformed ρ ; CI, 98.3% confidence interval for EVC, VOTC and LOTC; 95% CI for CM; 98.8% CI for size, elongation, compactness, and color. Highlighted in bold are the significant results; $\alpha = 0.017$ for EVC, VOTC, and LOTC; $\alpha = 0.05$ for CM; $\alpha = 0.0125$ for size, elongation, compactness, and color; n.s.= nonsignificant results with $P > 0.250$.

Nineteen-month-olds. The group-averaged DLT-RDM (Fig. 2A) showed an adult-like organization, as reflected by significant correlations with the composite-RDM, and the RDMs derived from the EVC, VOTC, and LOTC (Fig. 3C; see statistics in Table 1). The vector-of-ROIs analysis showed that the DLT-RDM was maximally correlated with object-related responses in early visual areas (V1 to V3) and fusiform gyrus ($P_s < 0.001$) (Fig. 3D).

Next, we asked which of the six categorical models underlying the composite-RDM, best represented the infants' DLT-RDMs. A stepwise linear regression ($\alpha_{\text{corrected}}$: 0.0083, two-tailed) showed an effect of the eight-category model [mean β = 0.090; 99.17% confidence interval (CI) [$100 \times (1 - 0.0083)$] = 0.026 to 0.155; $t(24) = 4.023$, $P < 0.001$; $d = 0.804$], the animacy model [mean β = 0.077; 99.17% CI = 0.014 to 0.139; $t(24) = 3.514$, $P = 0.002$; $d = 0.702$], and the humanness model [mean β = 0.133; 99.17% CI = 0.009 to 0.256; $t(24) = 3.091$, $P = 0.005$; $d = 0.618$] (for all other regressors: $P_s > 0.07$) (*SI Appendix*, Table S1).

In another analysis, we considered another measure of categorization: for the six categorization models, we tested whether average between-category DLTs were higher than average within-category DLTs ($\alpha_{\text{corrected}}$: 0.0083, one-tailed). Confirming the results above, we found that this was the case for the eight-category model [mean_{difference} = -0.135; 99.17% CI = $-\infty$ to -0.091; $t(24) = -7.967$, $P < 0.0001$; $d = 1.588$], the animacy model (mean_{difference} = -0.095; 99.17% CI = $-\infty$ to -0.041; $t(24) = -4.481$, $P < 0.0001$; $d = 0.896$) and the humanness model [mean_{difference} = -0.158; 99.17% CI = $-\infty$ to -0.060; $t(24) = -4.160$, $P < 0.001$; $d = 0.832$], but not for the other models (*SI Appendix*, Table S2).

Following evidence of categorization based on the eight-category model, we asked which of the eight categories the infants could indeed represent. Separately for each of the eight categories, we tested whether average within-category DLTs were lower than average between-category DLTs (t tests; $\alpha_{\text{corrected}}$: 0.0063, one-tailed). We found that, in addition to animates, inanimates, humans, and nonhumans, 19-mo-olds showed an ability to represent the subordinate categories of human bodies [mean_{difference} = -0.195; 99.37% CI = $-\infty$ to -0.060; $t(24) = -3.902$, $P < 0.001$, $d = 0.780$], nonhuman bodies [mean_{difference} = -0.160; 99.37% CI = $-\infty$ to -0.014; $t(24) = -2.956$, $P = 0.004$, $d = 0.603$], nonhuman faces [mean_{difference} = -0.216; 99.37% CI = $-\infty$ to -0.094; $t(24) = -4.763$, $P < 0.0001$, $d = 0.953$], and natural-small objects [mean_{difference} = -0.179; 99.37% CI = $-\infty$ to -0.076; $t(24) = -4.727$, $P < 0.0001$, $d = 0.986$] (all other $P_s > 0.024$).

Difference between images in color, size, elongation, or compactness did not account for the infants' looking behavior (Table 1), suggesting priority of categorical information, over more general physical differences, in the processing of visual objects at 19 mo.

Finally, we assessed possible preferences, considering the mean looking times (MLTs) toward each category, averaged across trials and subjects. A one-way repeated-measures ANOVA showed an effect of Category [$F(7, 168) = 16.259$, $P < 0.0001$; $\eta^2 = 0.790$], which reflected a preference (i.e., longer looking times) for animate (mean = 2.103 s \pm 0.344) over inanimate categories [mean = 1.598 s

± 0.247 ; $\text{mean}_{\text{difference}} = 0.505$; 95% CI = 0.367 to 0.643; $t(24) = 7.551$, $P < 0.0001$; $d = 1.509$], and for nonhuman animals (mean = 2.317 s \pm 0.463) over humans [mean = 1.889 s \pm 0.452; $\text{mean}_{\text{difference}} = -0.428$; 95% CI = -0.676 to -0.179; $t(24) = -3.552$, $P = 0.002$; $d = 0.711$] (see *SI Appendix, Supplementary Results* for details). The matrix representing t values for each pairwise comparison of MLTs remarkably replicated the structure of the DLT-RDM (Fig. 2A; see *SI Appendix, Table S3* for t and P values), showing categorization and discrimination based on animacy and humanness.

Ten-mo-olds. The looking behavior of 10-mo-olds (Fig. 2B) was significantly correlated with the composite-RDM of adult categorization as well as with fMRI-based RDMs reflecting object-related responses in selective aspects of the adults' visual ventral stream. Further analyses showed that objects were principally categorized by animacy.

More precisely, although correlations of the infants' DLT-RDM with activations in the broad ROIs (EVC, VOTC, and LOTC) did not reach the significance level (Fig. 3C and Table 1), the vector-of-ROIs analysis showed correlation with RDMs derived from the early visual cortex (V1) and fusiform gyrus ($P_s < 0.001$) (Fig. 3D). The DLT-RDMs also correlated with the synthetic composite-RDM. Infant's behavior was not explained by differences in visual feature, such as color, size, elongation, or compactness (Table 1).

Which of the six models underlying the composite-RDM best represented the infants' behavior? A stepwise linear regression showed correlation of the infants' DLT-RDM with the animacy model only [$\alpha_{\text{corrected}}: 0.0083$, two-tailed; mean $\beta = 0.059$; 99.17% CI = 0.002 to 0.117; $t(23) = 2.981$, $P = 0.007$; $d = 0.608$; for all other regressors: $P_s > 0.036$] (*SI Appendix, Table S1*). Consistent with this finding, average within-category DLTs were significantly lower than average between-category DLTs for the animacy model [$\text{mean}_{\text{difference}} = -0.061$; 99.17% CI = $-\infty$ to -0.007; $t(23) = -2.919$, $P = 0.004$; $d = 0.592$; $\alpha_{\text{corrected}}: 0.0083$, one-tailed], but not for the other models (*SI Appendix, Table S2*).

The analysis of the MLTs revealed an effect of Category [one-way repeated-measures ANOVA: $F(7, 161) = 9.422$, $P < 0.0001$; $\eta^2 = 0.808$], reflecting preference (i.e., longer looking times) for animate (mean = 2.044 s \pm 0.316) over inanimate categories [mean = 1.601 s \pm 0.263; $\text{m}_{\text{difference}} = 0.443$; 95% CI = 0.310 to 0.576; $t(23) = 6.901$, $P < 0.0001$; $d = 1.409$] (*SI Appendix, Supplementary Results 3*). Relationships between categories computed on the MLTs replicated the structure of the DLT-RDM, showing that 10-mo-olds categorized objects based on animacy, with a preference for animate objects (Fig. 2B and *SI Appendix, Table S3*).

Four-mo-olds. Unlike older infants, 4-mo-olds showed no evidence of categorization; they looked longer at human faces and big-inanimate objects or, otherwise, at the larger, less elongated and more compact of two images on the screen.

More precisely, the infants' DLT-RDM did not match the organization of object-related information in any broadly defined ROIs of the adults' visual cortex (Fig. 3C and Table 1), or in any smaller partition of the visual ventral stream (vector-of-ROIs analysis) (Fig. 3D). No correlation was found with the composite model of adult categorization (Table 1) or with any of the six underlying models (stepwise linear regression: all $t_s < 1$, n.s.) (*SI Appendix*, Table S1). For none of the categorization models were DLTs larger for between-category than within-category comparisons ($P_s > 0.032$) (*SI Appendix*, Table S2). Instead, infants' DLT-RDM correlated positively with the RDMs based on image size and compactness, and negatively with the RDM based on elongation (Table 1). No correlation was found with the color model.

A one-way repeated-measures ANOVA on the MLTs showed an effect of Category [$F(7, 161) = 22.970$; $P < 0.0001$; $\eta^2 = 0.802$], which was driven by a preference for human faces over all other categories ($\alpha_{\text{corrected}}: 0.0018$; all $P_s < 0.001$) (Fig. 2C and *SI Appendix*, Table S3). Within the inanimate categories, infants looked longer at big over small objects, whether artificial or natural ($P_s < 0.0001$). We note that the two most preferred categories (human faces and big-inanimate) were those with the largest image size ($>15,000$ pixels), the least elongated, and among the most compact shape (*SI Appendix, Supplementary Results 4* and Fig. S2). Thus, size, elongation, and compactness, rather than object identity, could explain object preferences in 4-mo-olds. In line with this, the MLTs computed for each image across subjects correlated positively with image size ($\rho = 0.515$, $P < 0.0001$) and compactness ($\rho = 0.397$, $P < 0.001$), and negatively with elongation ($\rho = 0.531$, $P < 0.0001$). That is, the larger the image, the less elongated, or the more compact the shape, the longer the looking time. Given this result, with a new stepwise linear regression, we reassessed the relationship of the DLT-RDM with the six categorical models, after removing the variance explained by size, elongation, and compactness. Yet, no model accounted for the remaining variance (all $P_s > 0.308$) (*SI Appendix, Supplementary Results 4* and Table S4). Likewise, we found no evidence of categorization comparing average within- and between-category DLTs (all $P_s > 0.136$) (*SI Appendix, Supplementary Results 4* and Table S5).

Comparison between groups. Categorical distinctions. The above analyses showed that categorization by animacy emerged by 10 mo, while categorization by humanness, and additional categories of the

eight-category model emerged by 19 mo. Additional between-subjects analyses confirmed the differences between age groups, with respect to object categorization by animacy, humanness, and by the eight-category model. In effect, for each of the three categorization models that appeared to change over time (i.e., across groups), we analyzed the variation of the mean difference between averaged between-categories vs. averaged within-category DLTs with a one-way ANOVA, including Age as between-subject factor (4 mo, 10 mo, 19 mo).

As for the animacy model, we found a trend for the effect of Age [$F(2, 70) = 2.752$; $P = 0.071$; $\eta^2 = 0.073$]. Pairwise comparisons showed that 4-mo-olds differed from 19-mo-olds [$\text{mean}_{4\text{-mo-olds}} = -0.022 \pm 0.121$ SD ; $\text{mean}_{19\text{-mo-olds}} = -0.095 \pm 0.106$; 95% CI = 0.009 to 0.139; $t(47) = 2.273$; $P = 0.028$; $d = 0.649$] but not from 10-mo-olds [$\text{mean}_{10\text{-mo-olds}} = -0.061 \pm 0.103$; 95% CI = -0.026 to 0.105; $t(46) = 1.225$; $P = 0.227$; $d = 0.354$]. Ten- and 19-mo-olds did not differ [95% CI = -0.026 to 0.094; $t(47) = 1.144$; $P = 0.258$; $d = 0.327$]. As for the humanness model, the effect of Age was significant [$F(2, 70) = 4.027$; $P = 0.022$; $\eta^2 = 0.103$]. Nineteen-mo-olds differed from both 4-mo-olds [$\text{mean}_{4\text{-mo-olds}} = -0.036 \pm 0.216$; $\text{mean}_{19\text{-mo-olds}} = -0.158 \pm 0.190$; 95% CI = 0.006 to 0.239; $t(47) = 2.109$; $P = 0.040$; $d = 0.602$] and 10-mo-olds [$\text{mean}_{10\text{-mo-olds}} = -0.003 \pm 0.202$; 95% CI = 0.043 to 0.268; $t(47) = 2.780$; $P = 0.008$; $d = 0.794$]. Four- and 10-mo-olds did not differ [95% CI = -0.155 to 0.088; $t(46) = -0.548$; $P = 0.586$; $d = 0.158$]. As for the eight-categories model, the effect of Age was significant [$F(2, 70) = 4.292$; $P = 0.018$; $\eta^2 = 0.109$]. Nineteen-mo-olds differed from both 4- [$\text{mean}_{4\text{-mo-olds}} = -0.048 \pm 0.120$; $\text{mean}_{19\text{-mo-olds}} = -0.135 \pm 0.085$; 95% CI = 0.028 to 0.147; $t(47) = 2.966$; $P = 0.005$; $d = 0.845$] and 10-mo-olds [$\text{mean}_{10\text{-mo-olds}} = -0.057 \pm 0.137$; 95% CI = 0.013 to 0.144; $t(47) = 2.426$; $P = 0.019$; $d = 0.690$]. Four- and 10-mo-olds did not differ [95% CI = -0.066 to 0.084; $t(46) = 0.243$; $P = 0.809$; $d = 0.070$].

In sum, categorization by humanness and by the so-called eight-categories model changed between 10 and 19 mo. Categorization by animacy differed between 4 and 19 mo. However, it did not differ significantly between 4 and 10 mo, although separate analyses for each group suggest a change in the representation of animate and inanimate categories between the two age groups. Given the robust evidence for categorization by animacy in 10-mo-olds, the absence of a difference between 4- and 10-mo-olds might suggest a latent categorization by animacy in the younger age group. Exp. 2 speaks to this question.

Lower-level dimensions. The above analyses showed that features such as size, elongation, and compactness contributed to driving the behavior of 4-mo-olds but not of 10- and 19-mo-olds. We assessed these differences across groups with three one-way ANOVAs, testing the effect of Age (4 mo,

10 mo, 19 mo) on the variation of correlation coefficients for the correlation between the DLT-RDMs and the RDMs for each of three visual dimensions (image size, elongation, and compactness). The correlation between looking behavior and image size changed over time [Effect of Age: $F(2, 70) = 21.259$; $P < 0.001$]. In particular, correlation was higher in 4-mo-olds than in 10-mo-olds [$\text{mean}_{4\text{-mo-olds}} = 0.315 \pm 0.173$; $\text{mean}_{10\text{-mo-olds}} = -0.068 \pm 0.208$; 95% CI = 0.136 to 0.359; $t(46) = 4.483$; $P < 0.001$; $d = 1.294$] and 19-mo-olds [$\text{mean}_{19\text{-mo-olds}} = -0.034 \pm 0.194$; 95% CI = 0.244 to 0.455; $t(47) = 6.649$; $P < 0.001$; $d = 1.903$]. Ten- and 19-mo-olds did not differ [95% CI = -0.014 to 0.217; $t(47) = 1.768$; $P = 0.084$; $d = 0.505$]. There was also a significant effect of Age for the correlation between elongation and looking behavior [$F(2, 70) = 22.180$; $P < 0.001$]. Correlation was higher in 4-mo-olds than in 10-mo-olds [$\text{mean}_{4\text{-mo-olds}} = -0.321 \pm 0.187$; $\text{mean}_{10\text{-mo-olds}} = -0.041 \pm 0.197$; 95% CI = -0.392 to -0.169; $t(46) = -5.051$; $P < 0.001$; $d = 1.458$] and 19-mo-olds [$\text{mean}_{19\text{-mo-olds}} = -0.010 \pm 0.151$; 95% CI = -0.409 to -0.214; $t(47) = -6.418$; $P < 0.001$; $d = 1.830$]. Ten- and 19-mo-olds did not differ [95% CI = -0.131 to 0.070; $t(47) = -0.614$; $P = 0.542$; $d = 0.175$]. Analogous results were found for the correlation between looking behavior and compactness [Effect of Age: $F(2, 70) = 16.694$; $P < 0.001$]: correlation was higher in 4-mo-olds than in 10-mo-olds [$\text{mean}_{4\text{-mo-olds}} = 0.208 \pm 0.155$; $\text{mean}_{10\text{-mo-olds}} = 0.054 \pm 0.190$; 95% CI = 0.054 to 0.255; $t(46) = 3.091$; $P = 0.003$; $d = 0.892$] and 19-mo-olds [$\text{mean}_{19\text{-mo-olds}} = -0.074 \pm 0.166$; 95% CI = 0.189 to 0.374; $t(47) = 6.144$; $P < 0.001$; $d = 1.757$], and higher in 19-mo-olds than in 10-mo-olds [95% CI = 0.025 to 0.230; $t(47) = 2.496$; $P = 0.016$; $d = 0.712$]. These results, together with the above correlation analysis, demonstrated that visual features of the stimuli such as image size, elongation, and compactness predicted the behavior of 4-mo-olds but not of older infants.

3.2.2. Experiment 2

In Exp. 1, 4-mo-olds showed no evidence of categorization, but a preference for human faces and big-inanimate objects, which might be explained by physical properties, such as image size, elongation, and compactness (i.e., a tendency to look at the larger/less elongated/more compact image on the screen). We asked whether a preference for certain physical properties might have overshadowed categorical effects. To this end, we tested a new group of 4-mo-olds ($n = 24$) with the same images of Exp. 1, but all matched for size (i.e., number of pixels) (Fig. 1C). Size, but not elongation and compactness, was modified because only the former can change without affecting object identity or recognizability. Results confirmed the preference for human faces and big-inanimate object, but also showed that, when size was no longer available to discriminate between two stimuli, 4-mo-olds showed categorization by animacy.

Specifically, the infants' DLT-RDMs (Fig. 2D) correlated with the composite-RDM, and the fMRI-based RDMs extracted from the anterior fusiform gyrus ($P < 0.001$) (Fig. 3D) in the vector-of-ROIs analysis (Table 1). There remained a significant negative correlation with the elongation model, a significant positive correlation with the compactness model, and no correlation with the color model (Table 1).

Of the six synthetic models that contributed to the composite-RDMs, infants' behavior was best represented by the animacy model [stepwise linear regression, $\alpha_{\text{corrected}}$: 0.0083, two-tailed; mean $\beta = 0.074$; 99.17% CI = 0.016 to 0.132; $t(23) = 3.697$; $P = 0.001$; $d = 0.755$; for all other regressors, $P_s > 0.12$] (*SI Appendix*, Table S1). The comparison between within-category and between-category DLTs confirmed this result, showing lower within-category than between-category DLTs for the animacy model [$\alpha_{\text{corrected}}$: 0.0083, one-tailed; mean_{difference} = -0.076; 99.17% CI = $-\infty$ to -0.021; $t(23) = -3.583$, $P < 0.001$; $d = 0.731$], but not for the other models (*SI Appendix*, Table S2).

A one-way repeated-measures ANOVA on the MLTs showed a significant effect of Category [$F(7, 161) = 59.466$; $P < 0.0001$; $\eta^2 = 0.869$], which reflected a preference for human faces over all other categories ($\alpha_{\text{corrected}}$: 0.0018; all $P_s < 0.0001$), for non-human faces over human or nonhuman bodies and for big- over small-inanimate objects (all $P_s < 0.001$) (*SI Appendix, Supplementary Results 3* and Table S3). Thus, the preference for human faces and big objects, which were the largest objects in Exp. 1, remained despite matching images for size. Moreover, the average MLTs for individual images were negatively correlated with elongation ($\rho = 0.337$, $P = 0.004$) and positively correlated with compactness ($\rho = 0.590$, $P < 0.001$), confirming the bias for the less elongated and more compact shapes. Given the last result, we reassessed the correlation of the DLT-RDM with the six categorical models, after removing the variance explained by elongation and compactness. Again, the animacy model was the only significant regressor [$\alpha_{\text{corrected}}$: 0.0083; mean $\beta = 0.055$; 99.17% CI = 0.005 to 0.106; $t(23) = 3.176$; $P = 0.004$; $d = 0.647$; for all other regressors, $P_s > 0.061$] (*SI Appendix*, Table S4). Categorization by animacy was confirmed by higher between- than within-category DLTs for the animacy model only [$\alpha_{\text{corrected}}$: 0.0083, one-tailed; mean_{difference} = -0.057; 99.17% CI = $-\infty$ to -0.009; $t(23) = 3.081$; $P = 0.003$; $d = 0.629$; for all other comparisons $P > 0.194$] (*SI Appendix, Supplementary Results 4* and Table S5).

Can face preference explain the categorization by animacy at 4 mo? The animate–inanimate distinction in a group that showed strong face preference could reflect the distinction between objects-with-face (preferred) vs. objects-without-face. In effect, infants looked longer at human faces but also

at non-human faces, compared to body stimuli (*SI Appendix*, Table S3). However, they also looked longer at human bodies, whose faces were not visible than nonhuman bodies, whose faces were visible. An additional analysis addressing the correlation between the DLT-RDM of 4-mo-olds and a synthetic model considering all the faces as one category, and all other stimuli as another category, only yielded a nonsignificant trend [mean $\beta = 0.055$; 99.17% CI = -0.027 to 0.137; $t(23) = 1.932$; $P = 0.066$; $d = 0.394$]. Thus, although it could contribute to it, the face preference cannot fully account for the animate–inanimate categorization in 4-mo-olds.

3.3. Discussion

Categorization is the mechanism through which the human mind makes sense of the environment by organizing the things of the world in categories. Categorization begins at a young age, with the ability to appreciate perceptual similarities between objects, and acquires refinement with knowledge and language acquisition. What type of real-world object categories infants can represent before developing a sizable lexicon and a rich system of knowledge about the world? We considered the hypothesis that the early stages of visual object categorization are guided by the same dimensions that structure object representations in the visual cortex of the primate brain.

Our findings demonstrate that early visual object categorization along the fundamental dimensions represented in the human visual cortex, is an incremental process with two milestones. The first, between 4 and 10 mo, establishes the transition from an exploration of the environment guided by general visual saliency to an organization that corresponds to the animate–inanimate categorical distinction; the second, between 10 and 19 mo, presents a spurt of visual object categories toward mature organization.

All the categorization effects that we observed analyzing looking-time differences between objects were paired with preference effects, as indexed by MLTs. That is, when infants showed categorizing objects by animacy, they also looked longer at animate than inanimate objects; when they showed categorizing objects by humanness, they also looked longer at nonhuman than human animals. The cooccurrence of preference and categorization suggests that the earliest visual categories to emerge in infancy are those that are important enough to give rise to a hierarchy of preferences.

However, we emphasize that categorization is not equal to preference. We operationalized categorization in terms of larger difference for between-category comparisons than within-category

comparisons. The systematic cooccurrence of effects in DLTs and in MLTs suggests that categorization effects were mainly carried by high between-categories DLTs. However, our approach can in principle detect categorization effects that are not paired with a systematic preference (i.e., categorical distinctions that do not yield consistent differences in the MLTs); for example, categorization by animacy could have been observed as long as within-category differences were lower than between-categories differences, and even if some infants preferred one category, and others, the other category yielding null group preference effect. Clearly, categorization effects are easier to detect, the lower the within-category differences (i.e., for rather homogenous object categories).

3.3.1. Within 4 mo

Four-mo-olds showed no evidence of categorization based on any of the categorical dimensions considered here, when image size allowed discriminating between two images on the screen (Exp. 1). Accordingly, their looking behavior did not match visual object representation in any sector of the adults' visual cortex. Infants' look, however, was not random. They looked longer at the larger image and the less elongated and more compact object on the screen, with size, elongation, and compactness differences predicting looking-time differences. Moreover, MLTs revealed preference for human faces and real-world size big (vs. small) inanimate objects.

The preference for human faces, extensively documented in very young infants (Farroni et al., 2005; M. H. Johnson et al., 1991; Valenza et al., 1996), has been explained by the detection of the characteristic eyes–mouth configuration, and iris-pupil-sclera contrast of human eyes (Farroni et al., 2005). However, if performance here reflected detection of those features, infants would have shown categorization of human faces, as all our human faces carried those features. Instead, preference for faces occurred without categorization (i.e., within-category DLTs were as high as between-categories DLTs). This suggests that, in processing two faces, infants focused on individual- rather than category-level features, possibly reflecting a propensity or need for individuation (i.e., for processing faces as individuals rather than category members) (Quinn & Eimas, 1998), which makes the differences between two faces as salient to the visual system, as the difference between a face and another object (Maurer et al., 2002).

The preference for real-world big (vs. small) objects also emerged without evidence of categorization. In this case, however, within-category DLTs comparable to between-category DLTs could reflect the visual heterogeneity of the category of big inanimate objects, which included landscapes (e.g., view of a lake), landmarks (e.g., building), and various large objects (e.g., washing-

machine and stool). This observation leaves open the possibility that categorization of big objects could be found for more homogeneous sets than the current one.

The effect of real-world object size at such a young age is unprecedented and open to multiple interpretations. Scholars are debating a proper characterization of the big-small object distinction, which might relate to differences in perceptual properties (e.g., texture, spatial frequency) (Long et al., 2018) and behavior-relevant properties of the objects (Magri et al., 2020). The current results add a new piece to the puzzle, showing early asymmetry in the attention toward big vs. small objects. In the adult brain, big objects, which typically function as landmarks, are represented in ventral aspects of the visual cortex, adjacent to place- and scene-specific areas. Small objects, which are by definition graspable, are represented in lateral aspects of the occipitotemporal cortex, also hosting areas for tool and action representation (Konkle & Oliva, 2012b). Areas of the scene- and place-specific network are functionally interconnected already in the first weeks of life (Kamps et al., 2020), and respond strongly to scenes in 4- to 6-mo-olds (Deen et al., 2017). In contrast, at 4 mo, infants are unable to grasp objects, showing immaturity of the networks that control hand movements and hand-object interaction. Interest in graspable objects increases during the first year of life (Libertus et al., 2013; C. Newman et al., 2001), as infants develop grasping skills (McCarty et al., 2001). Consistent with this trajectory, we found that by 10 mo, the preference for big over small objects had disappeared. Thus, different developmental trajectories of different networks in the visual cortex might contribute to object distinctions captured by the effect of real-world size.

3.3.2. From 4 to 10 mo

When size was no longer available to discriminate between two images on the screen (Exp. 2), 4-mo-olds continued to show a preference for human (as well as nonhuman) faces and big (vs. small) objects, but they also showed categorization by animacy. Thus, categorization by animacy was functional at 4 mo but was overshadowed by physical features, such as size, making an object more visible, independently from the category. By 10 mo, infants showed an ability to overcome the importance of low-level visual features in favor of categorical information: categorization by animacy emerged despite differences in image size. Moreover, by 10 mo, the preference for human faces and big (vs. small) objects had given way to interest in the broader category of animate entities.

Thus, the looking behavior of both 10-mo-olds in Exp. 1 and 4-mo-olds in Exp. 2 revealed categorization by animacy and matched the cortical organization of object-related information recorded

from anterior (temporal) aspects of the visual ventral stream in adults. Yet, a change between 4 and 10 mo happens as infants move from prioritizing image size to prioritizing categorical information.

Animacy is the earliest categorical distinction of visual objects in infancy. This implies that representation of animate entities is not an extension of the representation of conspecifics (Bonatti et al., 2002; Quinn & Eimas, 1998). Infants would rather start with a broad, underspecified representation of what animates look like, which might function as a coarse “life detector” to identify conspecifics as well as predators and preys (Vallortigara et al., 2005). The animate–inanimate distinction would thus lay the foundation for subtler categorical distinctions, and possibly sets conditions for domain-specific processes of naive psychology (Elizabeth S. Spelke & Amy E. Skerry, 2013; A. Woodward, 1998) vs. naive physics (Saxe et al., 2005; Spelke, 1994).

3.3.3. From 10 to 19 mo

With the second developmental change between 10 and 19 mo, infants showed an ability to represent the categories animate and inanimate, but also human, nonhuman, human bodies, nonhuman bodies, and nonhuman faces. The spurt of categories by 19 mo represents another step toward the model of mature visual object representation addressed here.

While in 4- and 10-mo-olds, categorization limited to two categories was associated with object-related responses in the most anterior aspects of the visual cortex, 19-mo-olds’ behavior correlated with object-related responses across the broad visual cortex of adults (from early visual cortex to ventral and lateral higher-level areas and from posterior to anterior regions along the ventral stream). This suggests that the ability to form new visual categories, from very general (e.g., animate–inanimate) to finer-grained (e.g., human vs. nonhuman bodies), involves the progressive recruitment of more and more feature spaces distributed over the visual cortex, and representing features with different complexity: as integration across regions (and feature spaces) increases, more and more categories can be represented. Promoter of this development, among other structural and functional maturation phenomena, could be the myelination of fiber tracts connecting distant areas (Dubois et al., 2014), which begins around 4 mo in the occipital lobe and continues later through the temporal lobe (Deoni et al., 2011). In the second year of life, categorization may further thrive with language development. Verbal labeling and communication of information about objects promote and shape the formation of new categories and, in some models, govern the transition from perceptual to conceptual categories (Bonatti et al., 2002; Dewar & Xu, 2007; LaTourrette & Waxman, 2020; Waxman & Markow, 1995; Westermann & Mareschal, 2014; Xu & Carey, 1996).

The developmental course of visual object categorization described here confirms and extends current knowledge of object categorization in infancy. Previous studies have shown that, by 5 mo, infants can represent the abstract categories of animate and inanimate, which constrain their ability to individuate objects (Bonatti et al., 2002; Surian & Caldi, 2010; Xu et al., 2004) and make inference about them. Thus, infants expect animate objects to be nonhollow (Setoh et al., 2013), intentional (Luo & Baillargeon, 2005; A. Woodward, 1998), to have beliefs (Baillargeon et al., 2010), social affiliation (Powell & Spelke, 2013, 2018), and morality (Dawkins et al., 2020; Hamlin, 2013). Inanimate objects are rather expected to be solid and obey the continuity principle (Spelke, 1994), to be moved by contact (Spelke, Phillips, et al., 1995), and lack intentionality (A. Woodward, 1998) and strong causal power (Muentener & Carey, 2010). In that body of research, infants identify animate entities primarily based on cues such as self-propelled motion, eyes, furry texture (Saxe et al., 2005; Setoh et al., 2013), and agentive/contingent behavior (Baillargeon et al., 2009; Deligianni et al., 2011; Markson & Spelke, 2006; G. E. Newman et al., 2010; Saxe et al., 2005; Setoh et al., 2013). Three- to 4-mo-olds can also learn to represent narrower, basic-level animate categories (e.g., dogs) after repeated exposure to various exemplars of a category to emphasize their visual similarity (Pauen & Peykarjou, 2021; Quinn et al., 1993, 2001; Quinn & Eimas, 1996). By contrasting two species, or basic-level categories [e.g., human vs. ape faces (Peykarjou et al., 2017); human body vs. horse, human body vs. cat (Quinn & Eimas, 1998; But see refs. Oakes et al., 1996; Pauen, 2000)], the human-nonhuman distinction has been observed even before 19 mo. Categorization of exemplars from homogeneous, basic-level categories (e.g., a few exemplars of canonical dogs) can rely on a few, very specific physical properties of the stimuli. In contrast, here, infants were faced with the harder task of extracting category-relevant information from a heterogeneous set of static visual features. That is, categorization required infants to recognize as members of the same (animate) category, human faces, zebras, fish, and parrots, on the one hand, and hammers, washing-machines, apples, and trees (inanimate objects), on the other hand, or to recognize that a human body and a human face fall in the same category, and a horse is closer to a fish than to a human body.

In studying when and how this task is achieved, the present study introduces two important advances. First, it shows that the infants' DLTs are a reliable measure of categorical similarity, with variations in DLTs reflecting variations in the similarity of image-computable features and categorical information (see refs. LaTourrette & Waxman, 2020 for converging evidence). Second, by using larger category boundaries than in previous studies, and without systematic manipulation of typical cues (e.g., self-propelled motion or agentive/contingent behavior for animacy), we have exposed in the infants'

looking behavior, the mechanism through which vision extracts category-relevant information from a large, heterogeneous set of features. The correlation between infants' looking behavior and fMRI data in adults suggests that infants can form categories using the same static visual information that yields categorical object representation in the visual ventral stream of primates. The so-formed visual categories could constitute another cue for early conceptual distinctions such as animate/inanimate and human/nonhuman. To test so, future research should study how infants label those categories, and what inferences they make about them: for example, do they infer from its static appearance that a crocodile is self-propelled, has intentions and beliefs?

Future studies should also address whether other categories can be captured in the infants' looking behavior by changing the category boundaries (e.g., more/less homogeneous categories), adding other real-world features (e.g., motion), or giving infants more time to explore the images. More, or finer-grained, categories could also be uncovered by going beyond the unidimensional characterization of the infants' looking behavior afforded by looking times (Kiat et al., 2021), or replicating the current methodology using neural correlates of infants' categorization. Finally, the exact nature of the category-relevant visual features that drove DLTs in infants remains to be studied. While we focused on visual features, as we targeted the model of object representation in the visual cortex, a role of other representational levels, mediated, for example, by language or semantic knowledge, remains to be tested.

3.3.4. Conclusions

We have shown that infants initiate their exploration of the visual world by giving priority to images that are more visible (i.e., the larger ones) and displaying preferences for faces and big objects. By 10 mo, they show the ability to learn that categorical information relevant to the animate–inanimate distinction is more important than general physical properties. Thus, the first act of visual object categorization divides the world into animate and inanimate entities. Other categories represented in the visual cortex emerge by 19 mo. As visual categories multiply, infants' behavior correlates with neural activity in ever-larger aspects of the adult visual cortex. Integration through growing connections within category-specific networks and between distant visual areas could be the driving force of this process. Increasing representation, and reliance on, category-relevant information in the first years of life may signal the coupling between seeing and thinking.

3.4. Materials and Methods

3.4.1. Eye-Tracking Study

Participants. The study involved 97 infants in total. Exp. 1 involved 24 infants of 4 mo (11 females; mean age: 4 mo, 15 d; range: 4:0 to 4:24), 24 infants of 10 mo (8 females; mean age: 10:26; range: 10:1 to 11:30), and 25 infants of 19 mo (11 females; mean age: 19:5; range: 18:1 to 20:1). Exp. 2 involved 24 infants of 4 mo (14 females; mean age: 4:17; range: 4:3 to 5:0). The sample size of 24 was arbitrarily chosen for the initial group of 19-mo-olds. Next, we verified that it was superior to the minimal sample size ($n = 18$) required to obtain the smallest categorical effect found in 19-mo-olds (human vs. nonhuman: $d_{\text{Cohen}} = 0.6182$, power = 0.80, $\alpha = 0.05$; GPower 3.1), and kept it constant across groups. We continued testing until we reached 24 participants per group. The last 19-mo-old infant had a twin; parents asked to test him too and we kept him in the sample. Fifty additional infants were tested in Exps. 1 and 2, but discarded (see *Analyses*). Infants were tested in the Babylab of the Institute of Cognitive Sciences Marc Jeannerod (Bron, France). Parents received travel reimbursement and gave informed consent before participation. The study was approved by the local ethics committee (CPP sud-est II).

Stimuli. We selected 72 total color photographs of isolated real-world objects from publicly available sets (Kiani et al., 2007) or from the internet. For Exp. 1, objects were superimposed on a gray background and scaled to fit a 350×350 pixels black frame. The final set of images consisted of nine exemplars for each of eight categories. Human faces were all female faces. For Exp. 2, all objects were resized to have the same number of pixels (54,135) without gray background (Fig. 1C).

Procedure. Infants sat on their parent's lap, ~60 cm away from a Tobii Eye-tracker T60XL screen, in a dark room. Parents were instructed to keep their eyes closed throughout the experiment. The experiment began after the calibration for eye-tracking and consisted of 36 trials. In a trial, two images were presented for 5 s on the left and right side of the screen, equally distant from central fixation (Fig. 1 B and C). Each infant saw a unique set of pairs including all 28 possible between-category combinations and eight within-category combinations. The experiment ended after 36 trials (~3 min), or because the infant expressed discomfort, or stopped looking at the screen. Stimulus presentation and data recording were controlled through PsyScopeX (psy.cns.sissa.it).

Analyses. On the eye-tracking screen, we defined two areas-of-interest overlapping with the locations of the two images. Areas-of-interest were two 350×350 -pixel squares in Exp. 1, and two masks

encompassing all nonbackground pixels in Exp. 2. For every trial, we computed the cumulative looking times in each area-of-interest.

Only trials with ≥ 1 s of look within the areas-of-interest were considered valid. For the analyses, we discarded infants with fewer than 27 valid trials (3 of 4 of total trials) or with a strong side bias (i.e., fixation on the same side for $>80\%$ of the experiment duration). In the final analysis of Exp. 1, 4-mo-olds contributed, on average, 35 ± 1 trials, 10-mo-olds, 34 ± 2 trials, and 19-mo-olds, 34 ± 2 trials. In Exp. 2, 4-mo-olds included in the final analysis contributed on average 29 ± 2 trials. Of all the infants tested in Exp. 1, exclusion criteria led us to discard 24 because of insufficient data (13 4-mo-olds, 7 10-mo-olds, and 4 19-mo-olds) and 1 4-mo-old because of a side bias. Twenty-five tested in Exp. 2 were discarded because of a side bias ($n = 2$) or insufficient data ($n = 23$). However, stimulus presentation differed between Exps. 1 and 2 (Fig. 1 B and C). In Exp. 1 areas-of-interest were fixed, squared areas delimited (and highlighted) by a black frame, which contained object and background. In Exp. 2, areas-of-interest overlapped with the object's contours without back-ground and frame. In sum, in Exp. 2, stimuli might have been less visually salient and, therefore, the exclusion criteria, more stringent, than in Exp. 1, yielding higher attrition rate. To address this, we carried out additional analyses adopting more lenient criteria in Exp. 2, to reach an attrition rate closer to that of Exp. 1. This analysis confirmed all of our results (*SI Appendix, Supplementary Results 5*).

For each infant, for each trial, one DLT was computed as the difference in the cumulative looking time (LT) between the right and the left area-of-interest divided by the sum of the two (i.e., the total time the infant attended to the areas of interest): $(LT_{\text{right}} - LT_{\text{left}})/(LT_{\text{right}} + LT_{\text{left}})$. Absolute and signed DLT values were entered in absolute and signed DLT-RDMs, respectively. Values on the diagonal (within-category DLTs) and off-diagonal values in one half of the DLT-RDM (between-category DLTs) were used for analysis.

Category effects. Separately for each experiment, for each group, we performed RSA to correlate the absolute DLT-RDMs with each of six categorization models, and the composite model of adult categorization, reflecting the average of the above six models. Each model defined an RDM, where cells had value of 0, 1, or 0.5 corresponding to within-category comparisons (i.e., lowest dissimilarity), between-category comparisons (i.e., highest dissimilarity), and comparisons irrelevant for a given categorization, respectively. For example, the humanness model had 0 for human–human (e.g., human face–human body) and nonhuman–nonhuman comparisons (e.g., cow face–elephant body), 1 for

human–nonhuman comparisons (e.g., human body–camel body), and 0.5 for irrelevant comparisons (e.g., artificial small object– natural large object).

First, we computed the correlation between the composite-RDM and the DLT-RDM of each infant. Individual Spearman correlation coefficients ρ for a group of infants were Fisher-transformed and tested against chance-level 0 (t test). Then, we performed a stepwise linear regression analysis for each infant, with the above six categorical models as regressors. For each regressor, the distribution of coefficients β in a group was compared against chance (t test). Categorization was further addressed by assessing whether, for each model, average within-category DLTs were higher than average between-category DLTs (t tests, one-tailed). All above analyses were computed considering the DLTs over the total 5-s trial duration.

Effects of general properties of the images. In Exp. 1, for each image, we computed a score for: 1) size (i.e., total number of pixels [350×350] minus number of background pixels); 2) shape-elongation (i.e., height-to-width ratio with ratio tending to 1 indicating lowest elongation); 3) shape-compactness, computed as the ratio between the area of the shape and the area of the disk with the same perimeter (values between 0 and 1, with 1 indicating maximal compactness); and 4) color (i.e., for the RGB format, the average of the mean values for red, green, and blue). Since each infant of a group saw different exemplars of a category, for each infant we created: an RDM representing signed size-differences for each pairwise comparison $[(\text{Size}_{\text{right-image}} - \text{Size}_{\text{left-image}})/(\text{Size}_{\text{right-image}} + \text{Size}_{\text{left-image}})]$; an RDM representing signed elongation-differences $[(\text{Elongation}_{\text{right-image}} - \text{Elongation}_{\text{left-image}})/(\text{Elongation}_{\text{right-image}} + \text{Elongation}_{\text{left-image}})]$; an RDM representing signed compactness-differences $[(\text{Compactness}_{\text{right-image}} - \text{Compactness}_{\text{left-image}})/(\text{Compactness}_{\text{right-image}} + \text{Compactness}_{\text{left-image}})]$; and an RDM representing color-differences in the form of Euclidean distance between the average color-value vector of two images. For each infant, we computed Spearman correlations between the DLT-RDM and each of the three RDMs. For each age group, individual correlation coefficients ρ were Fisher-transformed and entered in a one-sample t test (chance-level 0). For the signed RDMs (size, elongation, and compactness), positive correlation values indicated longer looking times toward larger/more elongated/compact objects.

Effects of size, elongation, and compactness on categorization. As size, elongation, and compactness correlated with the 4-mo' DLTs in Exp. 1, and elongation and compactness correlated with 4-mo' DLT in Exp. 2, we reassessed the effects of categorization after removing the variance explained by those physical properties of the images (*SI Appendix, Supplementary Results 4*). We performed a stepwise

linear regression on the signed DLT-RDM of 4-mo, with size RDM, elongation RDM, and compactness RDM as regressors. Next, we performed the stepwise linear regression analysis on the absolute values of the residual matrices R [$R = |\text{signed DLT-RDM} - \beta_{\text{size}} \text{size-RDM} - \beta_{\text{elongation}} \text{elongation-RDM} - \beta_{\text{compactness}} \text{compactness-RDM}|$].

Analysis of MLT. For each group, we computed the MLT toward each category. Differences across categories were analyzed with a one-way repeated-measures ANOVA and followed up with pairwise t tests.

3.4.2. fMRI Study on Adults

Participants. Fifteen participants took part in the fMRI study (eight females; mean age: $24.9 \text{ y} \pm 3.6$ SD). All had normal or corrected-to-normal vision, were screened for contraindications to fMRI, gave informed consent before participation, and were paid for their time. The local ethics committee (Comite de Protection des Personnes Sud Est V) approved the study.

Stimuli, procedures, and analyses. The fMRI study involved: 1) a main experiment to record neural responses to the same 72 object-stimuli shown to infants, and 2) a functional localizer session, used to define the whole visual ventral stream (from V1 to the fusiform gyrus) and three broad ROIs within the visual cortex (bilateral EVC, VOTC, and LOTC). fMRI data were used to create models of visual object categorization (i.e., RDM) based on the neural activity patterns evoked by the eight categories within each ROI, and, to capture more local effects of visual categorization, within each of the 38 consecutive slices along the antero-posterior axis that forms the visual ventral stream. RSA was used to test the correlation between the RDM extracted from each ROI and each slice of the ventral stream, and the DLT-RDM of each individual infant. See *SI Appendix, Supplementary Methods* for a detailed description of procedures and analysis.

Data Availability. Stimuli, group map fMRI data, eye-tracking data, and code for the main analyses have been deposited in the Open Science Framework repository created for this project (https://osf.io/6rm7a/?view_only=dcd418e45e074e379edee09ba36840be).

4. The roles of experience and maturation in the development of categorical saliency (study 2)

This chapter is a direct continuation of the previous study. The aim was to study the first stage of development previously describe, from 4 to 10 months of age. The first step was to better grasp the course of events between 4 and 10 months of age, studying 6 and 8 months. Then, the question was about understanding whether the brain maturation limits the influence and saliency of category membership in infants' behavior, or whether the experience drives category membership's influence on infants' behavior. The supplementary materials of this article can be find in Chapter 9, after the bibliography.

4.1. Introduction

Chapter 3 revealed a major developmental milestone in the saliency of category membership in infants' behavior. Indeed, while 4-month-old infants' behavior was guided by non-categorical features (i.e., features that are not eliciting the representation of a category; physical size on the retina, elongation and compactness), 10- and 19-month-old infants spontaneously relied on category membership, especially the animate-inanimate distinction, when exploring objects. Yet, we showed that 4-month-old infants already represented objects as animate or inanimate; but this representation was overshadowed by non-categorical features (Spriet et al., 2022). Thus, between 4 and 10 months of age, a major developmental step consists in a modification of the respective weights given to category membership (animate-inanimate) and non-categorical features, the former's weight increasing and the later decreasing. In the current chapter, we first tested 6- and 8-month-olds in order to inform the timeline of the observed development. Second, we ask what is the driving force of this transition: spontaneous maturation of the brain, or visual experience? We test preterm infants, whose visual experience is dictated by their chronological age, but their brain maturation depends on their corrected age, i.e., the age they would have is they were born at term.

Between 4 and 10 months of age, infants are developing a lot. They interact a lot more with objects and people in their environment at 10 months compare to 4 months: they can stand sited around 6 months, start to sit by themselves and crawl around 8 months of age (Malina, 2004; Marcinowski et al., 2019), reach and grasp smaller, more convenient objects for them to hold, around 6 months of age (Libertus et al., 2013), a capacity that still develops and improves around 8 months (Fagard et al., 2009;

C. Newman et al., 2001). They also demonstrate joint attention, gaze and point following by 7 to 8 months of age (Carpenter et al., 1998; Striano & Bertin, 2005). The development of motor skill, such as the ability to sit, as well as infants' social skills, change infants' interaction with objects in their environment, and thus their possibilities to learn about things in their environment (Libertus & Hauf, 2017; Soska & Adolph, 2014). This might partly explain the transition in the development between 4 and 10 months of age. Older infants, when exploring their environment, rely more on category membership than younger ones, as with the development of their motor and social skills, they are now able to interact with even more objects than before. In addition, infants' visual short-term memory is also improving between 6 and 8 months of age, as 6-month-old infants can only retain information (e.g., color) for one object presented in isolation, while 8-month-old infants are able to retain this information even if the object is presented in a multiple array (Kwon et al., 2014; Oakes et al., 2011). By 7 months, binding the color of an object to its location seems to be facilitated (Oakes et al., 2006, 2009). Another big development is the acceleration of visuo-attentional processes between 5 and 8 months of age (Hochmann & Kouider, 2022). Those last developments might in part explain a development of visual categorization ability, as older infants are now faster in extracting visual features of objects, better at binding those features to objects, and better at remembering them. Finally, one more developmental process has the potential to account for the switch in the reliability of category membership in infants' behavior: the brain myelination and maturation. The myelination of the occipital and parietal lobes starts in the 4th to 6th month of life, but only at 11 months of age the occipital lobe is at 50% of its complete myelination (Deoni et al., 2011). All those developmental processes can account for the first milestone described between 4 and 10 months of age, and refining the dynamic of the two steps of this milestone and the age of the developmental switch(es) will help us to understand what is happening at this stage of development. Thus, we extended the previous study, testing 6- and 8-month-old infants' reliance on category membership, especially diagnostic of the animate-inanimate categorization. To anticipate our findings, 8 months of age appears to be an age of transition, where infants both rely on category membership and non-categorical features.

After having identified this stage of development, we explored the respective role of the spontaneous brain maturation and of the visual experience. Both maturation and visual experience could theoretically account for the development of the saliency of the animate-inanimate categorization. On the one hand, maturation in the form of myelination of the occipital lobe, allows better and faster encoding of visual features and thus more efficient visual categorization. On the other hand, visual experience with objects and people is the opportunity for statistical learning of visual features that

concur in discriminating animate and inanimate objects. To explore the respective roles of brain maturation and visual experience on the saliency of category membership in infants' behavior, we tested healthy preterm infants. Myelination of preterm infants' brain largely depends on their corrected age, meaning the age they would have if they were born full-term (Tsuneishi & Casaer, 2007). Their visual experience, and more generally their sensory experience, in contrast, correlates with their chronological age (time since birth). Preterm infants' development is sometimes accelerated with respect to their corrected age, suggesting a major role of early experience in development (Caskey et al., 2011; Gonzalez-Gomez & Nazzi, 2012). Visual acuity, for instance, is equivalent in preterm and full-term 8-month-old infants (Van Hof-Van Duin & Mohn, 1986), thus higher than the acuity in infants of the same corrected age. Previous studies have used this approach, concluding that brain maturation plays a major role in the development of luminance sensitivity (Bosworth & Dobkins, 2009), language discrimination (Peña et al., 2010) but also, perhaps surprisingly, in the case of perceptual narrowing of phoneme discrimination (Peña et al., 2012). In contrast, visual experience plays the major role in the development of visual acuity (Van Hof-Van Duin & Mohn, 1986), sensitivity to contrast (Bosworth & Dobkins, 2009), binocular vision (Jandó et al., 2012) and gaze following (Peña et al., 2014).

We tested healthy preterm infants at 8 months of chronological age, whose corrected ages were about 6 months. This way, preterm infants had a brain maturation similar to the one of 6-month-old full-term infants, but a visual exposure as long as 8-month-old full-term infants. We found that preterm infants relied on visual features diagnostic of animate and inanimate categories, just like 8-month-old full-term infants. Altogether, the results suggested a major role of infants' experience in the use of category membership for exploring the world.

4.2. Materials and Methods

4.2.1. Participants

The study involved 96 infants in total. Experiment 1 involved 48 infants of 6 months in 2 different groups; one group of 24 infants (12 female; mean age: 6 months, 19 days; range: 6:4 to 6:30) was tested as 4-, 10- and 19-month-old infants were tested in the previous chapter (Experience 1, Chapter 3); the second group of 24 infants (12 female; mean age: 6 months, 15 days; range: 5:29 to 6:27) were tested as 4-month-old infants in Experiment 2 of the previous chapter. Experiment 2 involved 24 infants of 8 months (15 female; mean age: 8 months, 15 days; range: 8:0 to 9:1).

Experiment 3 involved 24 preterm infants of 8 months of chronological age (15 female; mean chronological age: 8 months, 17 days; mean corrected age: 6 months, 13 days; range of chronological age: 8:1 to 8:30; range of corrected age: 5:22 to 7:19). We are also acquiring data from preterm infants of 10 months of chronological age, but as I am writing this thesis, the group is not complete yet, thus I will not report the results here. The sample size of 24 was chosen based on previous experiment using the same methodology (Spriet et al., 2022). Thirty-six additional infants were tested in Experiments 1, 2 and 3, but discarded from the analyses, according to the same inclusion criteria as in the previous chapter.

4.2.2. Stimuli

The stimuli, procedure and analyses were the same as described in the previous chapter (Spriet et al., 2022), except that analysis focus on the animate-inanimate model (see Supplementary Results for the same analysis).

4.2.3. Analyses

In the final analysis of Exp. 1, 6-month-olds contributed, on average, 35 ± 2 trials and 8-month-olds, 35 ± 1 trials. In Exp. 2, 6-month-olds contributed on average to 29 ± 2 trials. In Exp. 3, 8-month-old preterm infants contributed on average 34 ± 2 trials. Exclusion criteria led us to discard 3 infants in Exp. 1 because of insufficient data (2 six-month-olds and 1 eight-month-old), 31 six-month-old infants in Exp. 2 and 4 eight-month-old preterm infants in Exp. 3. Stimulus presentation differed between Exps. 1-3 and 2, where stimuli might have been less visually salient in Exp. 2 compare to Exps. 1 and 3. This can explain the higher number of exclusion in Exp. 2 as compare to other Exps, as it was already the case in the previous chapter.

4.3. Results

4.3.1. Experiment 1

Two groups ($n = 48$, 24 per group) of infants aged 6 months saw 36 pairs of images, in the same design than in Chapter 3. Images in Experiment 1a were the same as in Experiment 1 of Chapter 3, while images in Experiment 1b were controlled for the size, as in Experiment 2 of Chapter 3. We mainly looked for the animacy model in infants' behavior, as well as the non-categorical features' influence (i.e., features that are not eliciting the representation of a category), as those are the influences that changes from 4 to 10 months of age.

Using representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008), we computed the relationship between RDMs based in infants' DLT (DLT-RDMs) and the model of animacy, as described in the previous chapter. We further explored the relationship between DLT-RDMs and the set of categorization models based on fMRI responses evoked in human adults in the previous chapter, with RDMs computed at each location along the antero-posterior axis of the visual ventral stream (i.e., *vector-of-ROIs* analysis). Finally, as infants' look was previously shown to be guided by physical properties of the stimulus, especially the size of the image, the elongation and the compactness, we also computed RDMs representing differences in size, elongation and compactness within pairs of images, keeping the signed values to have information about the image that was the larger, more compact or more elongated. We used those RDMs to study the implication of each physical property cited above as well as the implication of colors in the looking time of infants, keeping the signed values' DLT to know which image infants would look more (e.g., the larger, more compact image). These physical properties should only be considered as proxies of physical properties of our stimuli, as other image properties could affect infants' behavior.

Experiment 1a. Six-month-olds showed no evidence of categorization with the non-controlled images; just like 4-month-olds in the previous chapter, they looked longer at human faces and big-inanimate objects, or at the larger, less elongated and more compact image on the screen.

More precisely, the infants' DLT-RDM did not match the organization of object-related information in any smaller partition of the visual ventral stream (*vector-of-ROIs* analysis; Fig. 2). No correlation was found with the animacy model either (Table 1). The comparison between within-category and between-category DLTs also failed in revealing smaller within-category DLTs compare to the between-category DLTs ($M_{\text{difference}} = -0.024 \pm 0.092$ SD; 95% CI = -inf - 0.008; $t(23) = -1.279$; $P = 0.107$; $d = 0.261$).

However, infants' DLT-RDM correlated positively with the RDMs base on image size and compactness, and negatively with the RDM based on elongation (Table 1). No correlation was found with the color model.

A one-way repeated-measures ANOVA on the MLTs showed an effect of Category ($F(7,161) = 20.440$; $P < 0.0001$), which was driven by a preference for human faces over all other categories ($\alpha_{\text{corrected}}: 0.0018$; all $P_s < 0.0001$; Supplementary Table 4). Infants also looked longer at big ($M = 1.727$ s ± 0.296) compare to small inanimate objects ($M = 1.466$ s ± 0.282 ; $M_{\text{difference}} = 0.261$; 95% CI = 0.116 - 0.406; $t(23) = 3.731$; $P = .001$; $d = 0.761$).

As human faces and large objects are the larger, less elongated and more compact images, size, elongation and/or compactness, rather than object identity, could thus explain object preferences in 6-month-olds. In line with this, the MLTs computed for each image across subjects, correlated positively with image size ($\rho = 0.335$, $P = 0.004$) and compactness ($\rho = 0.397$, $P < 0.001$), and negatively with elongation ($\rho = -0.238$, $P = 0.044$). That is, the larger the image, the less elongated or the more compact the shape, the longer infant looked at it. Given this result, with a new stepwise linear regression, we reassessed the relationship of the DLT-RDM with the six categorical models, after removing the variance explained by size, elongation and compactness. Yet, no model accounted for the remaining variance (all P s > 0.02 ; Supplementary Results 4; Supplementary Table 5). Likewise, we found no evidence of categorization comparing average within- and between-category DLTs (all P s > 0.09 ; Supplementary Results 4; Supplementary Table 6).

Table 1. Results of representational similarity analysis reflecting relationships between the infants' DLT-RDMs and the models of visual categorization in adults, the model of animacy, and the RDMs based on size, compactness, elongation and color differences.

Exp.	Age	Model	mean ρ (<i>SD</i>)	CI (min – max)	t (df)	P	Cohen's d
1a	6 m	Animacy	.052 (.172)	-.021 – .125	1.485 (23)	.151	.302
		Size	.169 (.178)	.070 – .267	4.650 (23)	<.001	.949
		Elongation	-.125 (.201)	-.236 – -.014	-3.044 (23)	.006	.622
		Compactness	.181 (.143)	.101 – .260	6.174 (23)	<.0001	1.266
		Color	.057 (.155)	-.029 – .143	1.793 (23)	.086	.368
1b	6 m	Animacy	.078 (.152)	.014 – .142	2.509 (23)	.020	.513
		Elongation	-.235 (.241)	-.369 – -.102	-4.792 (23)	<.0001	.975
		Compactness	.378 (.228)	.252 – .505	8.120 (23)	<.0001	1.658
		Color	-.034 (.171)	-.128 – .061	-.965 (23)	<i>n.s.</i>	.199
2	8 m	Animacy	.188 (.178)	.113 – .263	5.170 (23)	<.0001	1.056
		Size	.144 (.177)	.045 – .242	3.963 (23)	<.001	.814
		Elongation	-.107 (.178)	-.205 – -.008	-2.937 (23)	.007	.601
		Compactness	.160 (.188)	.056 – .264	4.180 (23)	<.001	.851
		Color	.025 (.170)	-.069 – .119	.718 (23)	<i>n.s.</i>	.147
3	Preterm 8 m	Animacy	.089 (.177)	.014 – .164	2.457 (23)	.022	.503
		Size	.164 (.207)	.050 – .278	3.890 (23)	<.001	.792
		Elongation	-.119 (.191)	-.224 – -.013	-3.054 (23)	.006	.623
		Compactness	.155 (.171)	.060 – .249	4.428 (23)	<.001	.906
		Color	.012 (.173)	-.083 – .108	.350 (23)	<i>n.s.</i>	.069

Note: Exp., experiment; Age is the chronological age; m, months; mean ρ are the fisher transformed ρ ; CI, confidence interval; Highlighted in bold are the significant results; $\alpha = .05$ for Animacy model; $\alpha = .0125$ for Size, Elongation, Compactness and color except for Exp. 2 where $\alpha = .017$; *n.s.* = non-significant results with $P > .250$.

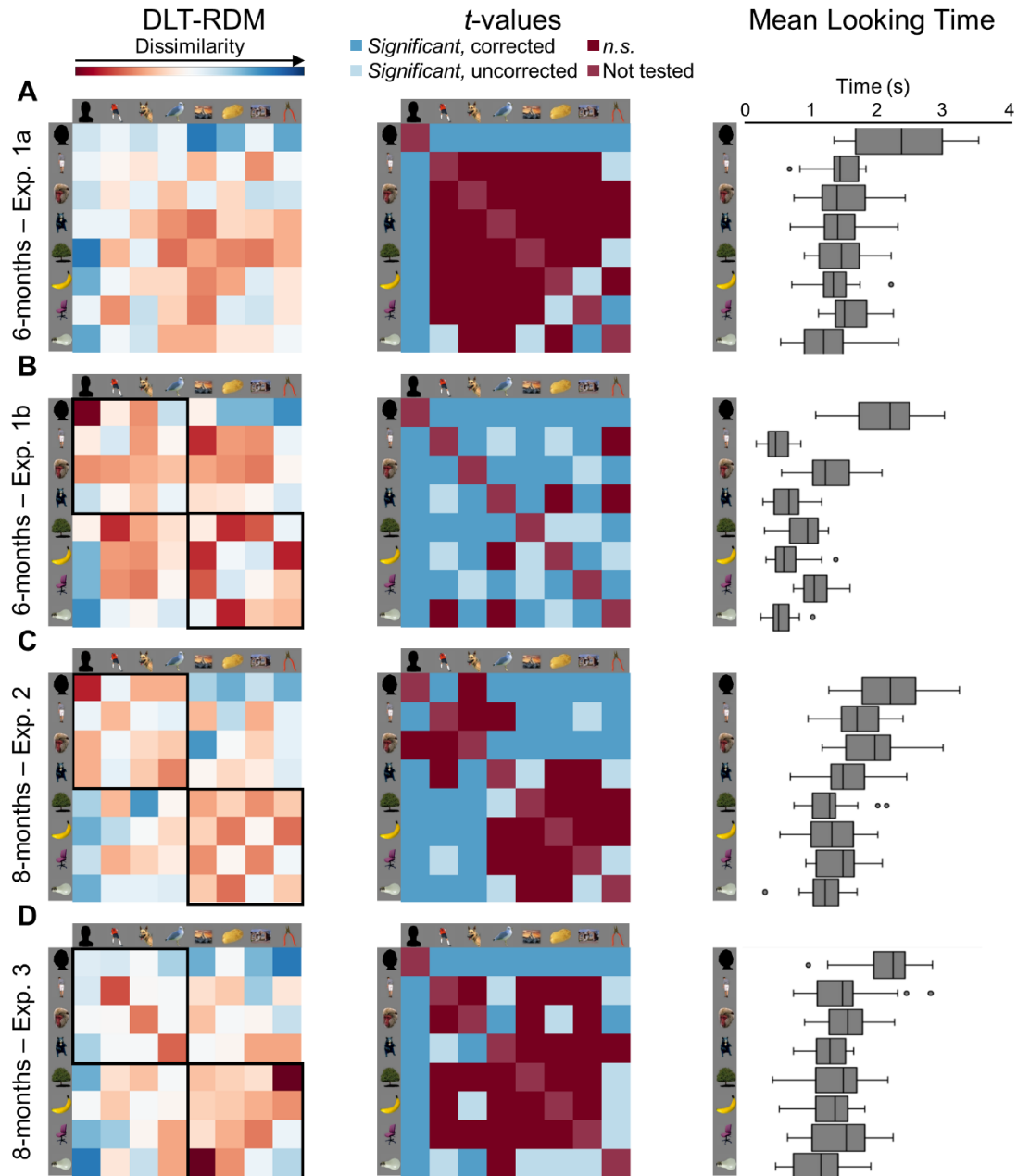


Fig. 1. Results of representational similarity analysis and of the pairwise comparisons of MLT between- and within-categories for each age group in Exps. 1, 2 and 3. **(Left)** Mean RDM reflecting dissimilarities between- and within-categories in terms of DLTs. Black squares in the RDMs highlight categorization by animacy in 6-month-olds in Exp. 1b **(B)**, 8-month-olds in Exp. 2 **(C)** and 8-month-old preterm infants in Exp. 3 **(D)**, and by the eight categories and animacy in 8-month-olds in Exp. 2 **(C)**. **(Center)** Matrix of t -values for each pairwise comparison between MLTs of the individual categories for 6-month-olds in Exp. 1a **(A)**, the 6-month-olds in Exp. 1b **(B)**, 8-month-olds in Exp. 2 **(C)** and 8-month-old preterm infants in Exp. 3 **(D)**. Squares in dark blue denote significant effects; squares in lighter blue denote effects that did not survive the multiple comparison correction (trends); red squares denote nonsignificant (n.s.) or nontested comparisons. **(Right)** Distribution of MLTs. Box-plots represent the minimum, the first quartile, the median, the third quartile and the maximum of the population distribution; outliers are denoted by dots.

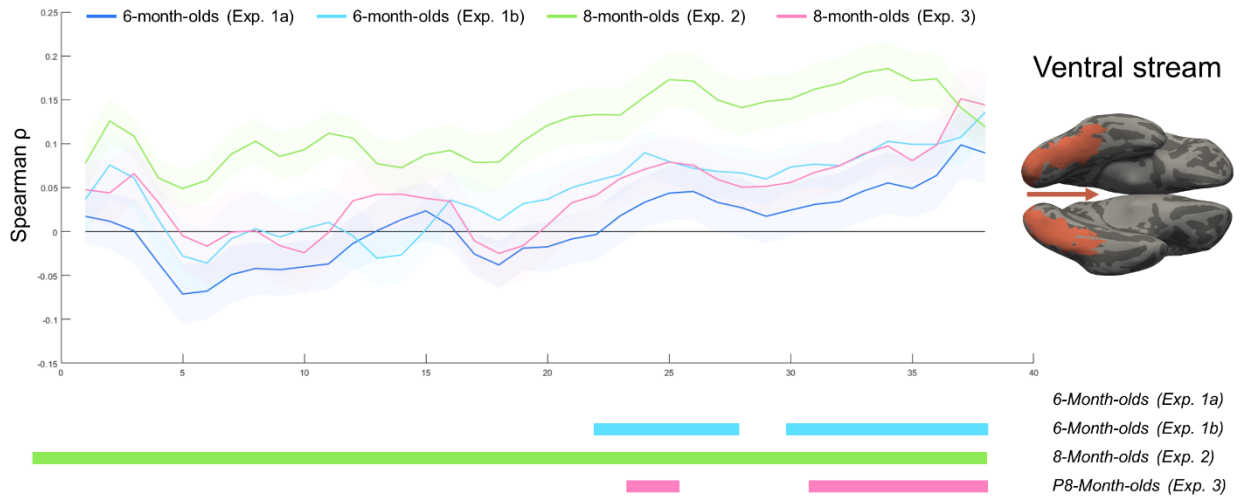


Fig. 2. Relationship between infants' looking behavior and visual object representation in the adults' visual cortex. **(A)** Results of the representational similarity analysis between the mean fMRI-based RDM (Spriet et al., 2022) in each ROI and the DLT-RDM of each infant in each age group of Exps. 1, 2 and 3. Box-plots represent the minimum, the first quartile, the median, the third quartile, and the maximum of the population distribution as well as outliers (dots); * $P < 0.017$; *** $P < 0.0003$. **(B)** Results of the representational similarity analysis between the infants' DLT-RDMs and the fMRI-based RDM (Spriet et al., 2022) derived from each partition along the ventral visual stream. Solid bars represent clusters with significant correlation (above 0) for each age group of Exps. 1, 2 and 3.

Experiment 1b. As 6-month-olds showed no evidence of categorization in Exp. 1a, but a preference for human faces and big-inanimate objects, just like 4-month-olds (Spriet et al., 2022), which might be explained by physical properties such as image size, elongation and compactness, we studied whether a preference for certain physical properties might have overshadowed categorical effects, and tested a new group of 6-month-olds ($n = 24$) with the same images, but matched for size (number of pixels). Only size was modified because it is the only feature of the three explored that can change without affecting object identity or recognizability. Results confirmed the preference for human faces and for big-inanimate objects, and, like for 4-month-olds (Chapter 3), revealed evidence for the animate-inanimate categorization.

Specifically, the infants' DLT-RDMs correlated with the animacy model (Table 1) as well as with the fMRI-based RDMs extracted from the anterior fusiform gyrus in the vector-of-ROIs analysis (Fig. 2). There remained a significant negative correlation with the elongation model, a significant positive correlation with the compactness model and no correlation with the color model (Table 1).

The correlation with the animacy model was confirmed by the comparison between within-category and between-category DLTs, showing that within animate and inanimate DLTs were

significantly smaller than between animate/inanimate DLTs ($M_{\text{difference}} = -0.041 \pm 0.095$ SD; 95% CI = -inf – -0.008; $t(23) = -2.139$; $P = 0.022$; $d = 0.432$). For other models, see Supplementary Results 1.

A one-way repeated measures ANOVA on the MLTs showed an effect of Category ($F(7,161) = 73.463$; $P < 0.0001$), which was driven by a preference for human faces over all other categories ($\alpha_{\text{corrected}}: 0.0018$; all $P_s < 0.0001$; Supplementary Table 4), for animate compare to inanimate (10 out of 16 comparisons are significant with $P_s < .001$; 3 additional comparisons would be significant without the multiple tests correction), for big objects compare to small ones, whether artificial or natural (3 out of 4 comparisons are significant with $P_s < .0001$; the remaining does not survive the multiple tests correction but would be significant without it, $P = .004$), for humans compare to nonhuman animals (3 out of 4 comparisons are significant with $P_s < .0001$; the remaining one would be without the multiple tests correction, $P = .014$) as well as for faces compare to bodies (all $P_s < .0001$). Infants looked indeed longer at big ($M = 1.117$ s ± 0.230) compare to small inanimate objects ($M = 0.716$ s ± 0.221 ; 95% CI = 0.296 – 0.507; $t(23) = 7.847$; $P < 0.0001$; $d = 1.598$), as well as at animate ($M = 1.295$ s ± 0.189) compare to inanimate ($M = 0.916$ s ± 0.188 ; 95% CI = 0.266 – 0.491; $t(23) = 6.947$; $P < 0.0001$; $d = 1.420$), human ($M = 1.490$ s ± 0.257) over nonhuman ($M = 1.100$ s ± 0.284 ; 95% CI = 0.227 – 0.553; $t(23) = 4.944$; $P < 0.0001$; $d = 1.010$), and faces ($M = 1.928$ s ± 0.422) over bodies ($M = 0.662$ s ± 0.175 ; 95% CI = 1.046 – 1.488; $t(23) = 11.857$; $P < 0.0001$; $d = 2.423$). Thus, the preference for human faces and big objects, which were the largest objects in Experiment 1, remained despite matching images for size. Moreover, the average MLTs for individual images were negatively correlated with elongation ($\rho = -0.407$, $P < 0.001$) and positively correlated with compactness ($\rho = 0.603$, $P < 0.0001$), confirming the bias for the less elongated and/or more compact shapes. Given the last result, we reassessed the correlation of the DLT-RDM with the six categorical models, after removing the variance explained by elongation and compactness. However, none of the six synthetic models explained infants' behavior (all $P_s > 0.07$; Supplementary Table 5), and none of the between-category and within-category comparisons reveal significant differences either (all $P_s > 0.19$; Supplementary Table 6).

4.3.2. Experiment 2

After having replicated 4-month-old infants' behavior with 6-month-old infants, we explored the spontaneous use of visual features diagnostic of animate and inanimate categories in 8-month-old infants. The group-averaged DLT-RDM showed an adult-like organization (Fig. 1), as shown the

vector-of-ROIs analysis, revealing that the DLT-RDM correlated all along the visual ventral stream, especially with the fusiform gyrus (where all P s < 0.001; Fig. 2). The significant correlation between infants' DLT-RDM and the model of Animacy (see statistics in Table 1; Fig. 1) also revealed the spontaneous guidance of 8-month-old infants' look by category membership, further demonstrate by the smaller within-category DLTs compare to the between-category DLTs for animate and inanimate categories ($M_{\text{difference}} = -0.111 \pm 0.098$ SD; 95% CI = -inf - -0.077; $t(23) = -5.556$; $P < 0.0001$; $d = 1.133$). For other models, see Supplementary Results 1.

Infants' DLT-RDM also correlated positively with the RDMs based on image size and compactness, and negatively with the RDM based on elongation (Table 1). No correlation was found with the color model. Given this result, with a new stepwise linear regression, we reassessed the relationship of the DLT-RDM with the six categorical models, after removing the variance explained by the size, compactness and elongation of images. This analysis confirmed the effect of the animacy model (Supplementary Results 4; Supplementary Table 5). Likewise, we found evidence of categorization according to the animacy model comparing average within- and between-category DLTs (Supplementary Results 4; Supplementary Table 6).

We then assessed possible preferences, considering the mean looking times (MLTs) toward each category, averaged across trials and subjects. We thus performed a one-way repeated-measures ANOVA, which showed an effect of Category ($F(7,161) = 18.248$; $P < 0.0001$), which reflected a preference (longer looking times) for human faces over almost all other categories except over nonhuman animal faces ($\alpha_{\text{corrected}}$: 0.0018; all P s < 0.001 except for the comparison with nonhuman animal faces where $P > 0.08$; Supplementary Table 4 for t - and P -values of all comparisons). It also reflected a preference for animate ($M = 2.066$ s \pm 0.317) over inanimate categories ($M = 1.483$ s \pm 0.234; $M_{\text{difference}} = 0.583$; 95% CI = 0.434 - 0.731; $t(23) = 8.120$; $P < 0.0001$; $d = 1.656$), as well as for faces ($M = 2.296$ s \pm 0.428) over bodies ($M = 1.836$ s \pm 0.328; $M_{\text{difference}} = 0.461$; 95% CI = 0.282 - 0.639; $t(23) = 5.331$; $P < 0.0001$; $d = 1.090$).

Finally, because 8-month-old infants' looking behavior is also driven by physical properties of the images, and because of the preference for at least human faces, which were the largest objects in Experiment 1, we assessed the correlation between the average MLTs for individual images and each physical properties influencing infants' behavior. The average MLTs were not correlated with the size ($\rho = 0.211$, $P = 0.075$) nor with elongation ($\rho = -0.072$, $P = 0.547$) but positively correlated with compactness ($\rho = 0.308$, $P = 0.009$), confirming the bias for the more compact shapes.

Comparison between groups

Categorical distinctions. Results showed that categorization by animacy model emerged by 8 months. Additional between-subject analyses confirmed the differences between age groups, with respect to object categorization by animacy. We analyzed the variation of the mean difference between averaged between-categories *vs.* averaged within-category DLTs with a two sample *t*-test, exploring the effect of Age (6 months and 8 months), and found a significant difference between the younger and the older groups of infants ($M_{6\text{-month-olds}} = -0.024 \pm 0.092$ SD; $M_{8\text{-month-olds}} = -0.111 \pm 0.098$; 95% CI = 0.032 – 0.143; $t(46) = 3.183$; $P = 0.003$; $d = 0.784$).

Additional analysis, including 4-, 10- and 19-month-old infants' data from Chapter 3 revealed an increase of category distinctiveness with age, i.e., an increase of the difference between within and between category comparisons, either because of a smaller dissimilarity within category and/or because of a higher dissimilarity between category (see Supplementary Results 6; Supplementary Fig. 3).

Lower-level dimensions. The above analyses also showed that features such as size, elongation and compactness contributed to driving the behavior of both 6- and 8-month-olds. We assessed whether a difference across groups exists considering those dimension, with a two-sample *t*-test, testing the effect of Age (6 months and 8 months) on the variation of correlation coefficients (fisher transformed) for the correlation between the DLT-RDMs and the RDMs for each of three visual dimensions (image size, elongation, and compactness).

The correlation between looking behavior and image size, elongation and compactness do not appear to change between 6 and 8 months of age. Pairwise comparisons showed no difference between 6- and 8-month-olds for image size ($M_{6\text{-month-olds}} = 0.169 \pm 0.178$ SD; $M_{8\text{-month-olds}} = 0.144 \pm 0.177$; 95% CI = -0.078 – 0.129; $t(46) = 0.493$; $P = 0.624$; $d = 0.093$), elongation ($M_{6\text{-month-olds}} = -0.125 \pm 0.201$ SD; $M_{8\text{-month-olds}} = -0.107 \pm 0.178$; 95% CI = -0.128 – 0.092; $t(46) = -0.329$; $P = 0.744$; $d = 0.070$) nor compactness ($M_{6\text{-month-olds}} = 0.181 \pm 0.143$ SD; $M_{8\text{-month-olds}} = 0.160 \pm 0.188$; 95% CI = -0.077 – 0.117; $t(46) = 0.421$; $P = 0.676$; $d = 0.083$).

4.3.3. Experiment 3

In Experiments 1-2, we found that both 6- and 8-month-olds explore their visual inputs relying on physical properties of the image such as the size, compactness and elongation, while only 8-month-old infants (and not 6-month-olds) also relied on the category membership, especially on the features

diagnostic of animate and inanimate categories. Exp. 1b further showed that, when controlling for the size of images, 6-month-olds demonstrated their capacity to represent the categories of animate and inanimate. In doing so, they behaved exactly like 4-month-old infants (Spriet et al., 2022). These data suggest two sequential developments: (1) between 6 and 8 months, the weight of categorical distinctions (animate-inanimate) increases; (2) between 8 and 10 months, the weight of non-categorical features decreases.

We now explore the respective roles of brain maturation and visual experience in the first development, by testing preterm infants at 8 months of chronological age, with corrected ages of 6 months. Thus, their brain maturation is similar to the one of 6-month-old full-term infants, but they have the same amount of visual exposure as 8-month-old full-term infants. If the brain maturation triggers the transition, 8-month-old preterm infants would then behave as 6-month-old full-term infants. However, if the experience drives infants' attention towards category membership, 8-month-old preterm infants would behave as 8-month-old full-term infants.

Just like 8-month-old full-term infants, 8-month-old preterm infants showed a clear evidence of animate-inanimate categorization; they also looked longer at human faces, or at the larger, less elongated and more compact image on the screen. Their behavior also compared to the organization found in adults' ventral stream. Indeed, 8-month-old preterm infants' behavior correlated with a small part of the adult ventral stream, although a smaller portion as compare to 8-month-old full-term infants.

More precisely, the infants' DLT-RDM matched a small partition of the visual ventral stream, in the anterior part of the fusiform gyrus (vector-of-ROIs analysis; Fig. 2). In addition, infants' DLT-RDM did correlate with the Animacy model (Table 1; Fig. 1). The correlation with the animacy model was confirmed by the comparison between within-category and between-category DLTs, showing that within animate and inanimate DLTs were significantly smaller than between animate/inanimate DLTs ($M_{\text{difference}} = -0.051 \pm 0.111$ SD; 95% CI = -inf - -0.012; $t(23) = -2.243$; $P = 0.017$; $d = 0.460$). For other models, see Supplementary Results 1.

As for 6- and 8-month-old full-term infants, 8-month-old preterm infants' DLT-RDM correlated positively with the RDMs base on image size and compactness, and negatively with the RDM based on elongation (Table 1). No correlation was found with the color model.

A one-way repeated-measures ANOVA on the MLTs showed an effect of Category ($F(7,161) = 17.873$; $P < 0.0001$), which was driven by a preference for human faces over all other categories ($\alpha_{\text{corrected}}: 0.0018$; all $P_s < 0.0001$; Supplementary Table 3).

Because Experiments 1-2 showed that size, elongation and/or compactness, rather than object identity, could indeed explain object preferences in 6- and 8-month-old full-term infants, we explored again the correlation between the MLTs computed for each image across subjects, with image size ($\rho = 0.276$, $P = 0.019$), compactness ($\rho = 0.381$, $P = 0.001$), and elongation ($\rho = -0.189$, $P = 0.111$), and found that both the size and elongation could account for infants' preferences. We thus reassessed the relationship of the DLT-RDM with the six categorical models, with a new stepwise linear regression, after removing the variance explained by size, elongation and compactness. This analysis revealed again that the animacy model explained infants' DLT variance (Supplementary Results 4; Supplementary Table 5). Likewise, we found categorization by animacy when comparing average within- and between-category DLTs (Supplementary Results 4; Supplementary Table 6).

As we were interested into comparing preterm infants' behavior to full-term infants' behavior, we assessed which of the two averaged full-term infants' behavioral matrices best represented the infants' DLT-RDMs, performing two correlations between 8-month-old preterm infants' DLT-RDM and the averaged DLT-RDM of 6-month-old infants in Exp. 1a and the averaged DLT-RDM of 8-month-old infants in Exp. 2. This analysis revealed that, while 8-month-old preterm infants' behavior mimicked 8-month-old full-term infants' behavior in their use of animate-inanimate features, but not so much 6-month-old full-term infants' behavior, 8-month-old preterm infants' behavior was corresponding to 6-month-old full-term infants' behavior ($M_{\text{fisherp}} = 0.110 \pm 0.216$ SD; 97.5% CI = 0.020 – 0.215; $t(23) = 2.493$; $P = 0.020$; $d = 0.509$) as well as with 8-month-old full-term infants' behavior ($M_{\text{fisherp}} = 0.103 \pm 0.234$ SD; 97.5% CI = -0.012 – 0.217; $t(23) = 2.145$; $P = 0.043$; $d = 0.438$).

4.4. Discussion

Following the previous chapter, we explored the transitional age at which the features important for the animate-inanimate categorization gain sufficient salience to guide infants' behavior in their visual exploration of objects. We further sought to understand the respective roles of the brain maturation and visual experience in this development. Our results revealed that infants started to spontaneously rely on category memberships at 8 months of age, though non-categorical visual features (i.e., features that are not eliciting the representation of a category) such as the physical properties of the images still influenced them, pointing to a milestone at 8 months in the visual categorization development. This transition was mainly driven by experience, as 8-month-old preterm infants demonstrated a behavior guided by animate-inanimate features as well.

First step of the visual categorization development. Younger infants did not rely on category membership when exploring objects. The representation of objects by category was present, but hindered by a greater reliance on other, non-categorical visual features. Indeed, 4- and 6-month-old infants looked longer at the bigger, more compact and less elongated objects in a pair, without paying attention to the category of the images. However, once objects were presented at the same size, controlling for some non-categorical visual features, infants showed evidence of categorizing images according to animacy.

By 8 months, infants' looks were spontaneously guided by category membership when exploring objects, especially those responsible for the animate-inanimate categorization. Their behavior was still influenced by non-categorical visual features as well: they also looked longer at the bigger, more compact and less elongated images. This pattern is supporting a two-stages development in the saliency of visual features diagnostic of a category in infants' behavior. First, between 6 and 8 months of age, infants start to pay more attention to category membership, and explore images according to both the non-categorical features *and* the features diagnostic of a category of the images. Second, between 8 and 10 months of age, infants attribute less importance to non-categorical features and only explore images based on their category membership. Eight-month-old infants' behavior also reflected the organization of object categorization observed in the adults' visual ventral stream, in both posterior and anterior regions. It was only the case in anterior regions for 6 month-olds, and only when image size was controlled.

Effect of brain maturation and experience. By studying preterm infants that have the same chronological age as 8-month-old full-term infants but a corrected age of 6 months, thus a brain maturation equivalent to that of 6-month-old full-term infants, we were able to evaluate the respective roles of brain maturation and visual experience on the saliency of features diagnostic of a category in infants' exploration of the world. Our results revealed that for the category membership to become salient enough to drive infants' behavior, infants need visual experience. Indeed, 8-month-old preterm infants' behavior paralleled the behavior of infants with the same chronological age: 8-month-old full-term infants. Both groups paid attention to non-categorical features but also categorized images according to animacy. Both groups' behaviors reflected the organization of object categorization as found in the adults' visual ventral stream.

The effect of categorization appears however stronger in full-term 8-month-olds than in preterm 8-month-olds, and the full-terms' behavior reflected the organization of a larger portion of the ventral

stream than preterms' behavior. Thus, even though our main analyses highlighted the role of visual experience, a subsidiary role for maturation is not to be excluded.

Importantly, we repeatedly showed, in 6-month-olds and in 4-month-olds (Chapter 3), that visual categorization is present in young infants. What develops between 6 and 8 months, as a consequence of visual experience, is not the categories per se, but rather their weight in guiding infants looking behavior. Below we begin a reflection about the type of visual experience and the learning mechanism that might account for the observed development.

Between 6 and 8 months of age, infants develop in many respect. They start to be able to seat around 6 months of age, but still depend on their parents reach the seating position. Only by 8 months do they start to actually seat by themselves, in addition to be able to crawl (Malina, 2004; Marcinowski et al., 2019). A major difference between 6 and 8 months of age seems to be the infants' independence in their interaction: they start to move by themselves, they can now better chose with what and when to interact with objects in their environment. They suddenly also exponentially increase their possibilities to learn about objects in their environment (Libertus & Hauf, 2017; Soska & Adolph, 2014). In addition to those motor developments, they also demonstrate an acceleration in the processing of visual information (Hochmann & Kouider, 2022) and developments of social skill, such as the beginning of joint attention, and interpreting referential gazes and pointing (Carpenter et al., 1998; Striano & Bertin, 2005).

The increasing weight of categorical information around 8 months may be related to the increased amount of varied interactions with objects, consequences of independent locomotion and increased social interactions. If so, the interpretation of the development of preterm infants observed in Experiment 3 depends on the respective roles of maturation and experience in the development of motor and social skills. Gaze following developed at 7 months of age, in both full-term and preterm infants (Peña et al., 2014), while prematurity had different effect on different motor skills, either shown as delayed compare to infants of the same corrected age, or shown to partly differ between infants of the same chronological age but not in a majority of preterm infants (Celik et al., 2018; see for review Fuentefria et al., 2017), pointing towards social skills as potential co-variates of the increase weight of categorical information in infants' behavior.

Another account for this switch comes from a more specific visual exposure. Indeed, while infants grow older and become more independent in their exploration of the world, they are still evolving in a specific environment. They are, in general, exposed to their caregivers, some example of

animals, as well as some examples of natural and artificial objects. While being exposed to all those objects, their visual system might become tuned toward the statistical regularities of those objects, and as their brain is already able to pick up those regularities, being sufficiently exposed to those objects leads to the saliency of category membership in driving infants' attention.

This question could be address with preterm infants, although it was not the aim of this study. Studies have shown that motor development in preterm infants was mainly delayed, not only as compare to their chronological, but also corrected age (Celik et al., 2018; see for review Fuentefria et al., 2017), although not a majority of preterm infants were showing delay in their motor development compare to their corrected age (Valentini et al., 2019). However, infants' gaze following appeared to be equivalent in preterm and full-term infants of the same chronological age (Peña et al., 2014), suggesting equivalent gaze control in preterm infants compare to their chronological age. Coupled with an adequate screening of visual exposition (e.g., reading book with many examples of animals), and a better inspection of the time spent in a neonatal care unit after being born, comparing preterm infants with and without social delays might inform further studies about the specific and aspecific role of visual experience and overall experience in the category membership's saliency.

Conclusion. We have shown that by 8 months of age, once infants were sufficiently exposed to their environment, features diagnostic of a category become salient enough to influence infants' looking behavior. Infants are still sensitive to the stimuli size, elongation and compactness, but they also give higher weight to their category membership, especially their animacy.

5. Perceptual acceleration enhances visual object categorization (study 3)

This chapter focuses on the animate-inanimate superordinate categorization, using a set of images selected to represent the big variability of within each category of animate and inanimate objects (photographs of objects superimposed on a gray background). It aims at examining the speed of categorization, in adults and infants in the first year of life.

5.1. Introduction

Visual categorization is at the basis of our reasoning. It allows us to start making sense of the environment in a glance, and it is already present early on in life (Kellman & Arterberry, 2007; Slater, 2002). Categorizing an object leads us to automatically retrieve information attached to this object, and permits to act accordingly (*e.g.*, fleeing a predator; eat an apple). The visual categorization is based on the accumulation of evidence, from encoding simple lower-level visual features such as the color, the spatial frequency, the orientation or the contrast of objects, that decay fast in time, to encoding higher-level visual features via the abstraction of the visual input, that decay more slower in time and lead to categorical representations, and to the classification and identification of the object (Groen et al., 2017). Previous studies have found that categorical brain response were only captured when adults report perceiving a face, even in objects stimuli that had no face but were arranged in a face configuration (*e.g.*, eggs and bacon in a pan), suggesting that processing visual information into categories elicits perceptual awareness in adults (Rekow, Baudouin, Brochard, et al., 2022; Retter et al., 2021). Perceiving visual categories thus reflects adults' ability to perceive the object, be aware of it, and think about it. Evidence for infants' capacity to categorize objects were already described at 4 months of age, especially for the human face categorization (de Heering & Rossion, 2015; Leleu et al., 2020; Rekow et al., 2021), as well as for larger categories, such as mammals and furniture (Peykarjou et al., 2023, 2024), or animate and inanimate objects (Spriet et al., 2022). The categorization of animate and inanimate objects reflect a fundamental step for reasoning developments as it may drive different kinds of inferences in infancy, referred to as naïve psychology and naïve physics (Carey, 2009; Spelke, 2022). That is, infants have different expectations and apply different principles (the principles of naïve psychology) when reasoning about animate agents that have goals (A. Woodward, 1998), beliefs (Kovács et al., 2010) and causal power (Adibpour & Hochmann, 2023; Muentener & Carey, 2010), and

inanimate objects, that are processed according to the laws of physics (naïve physics; Carey, 2009; Spelke, 2022).

The processing of visual information begins at birth but gets faster and faster throughout development, and especially accelerates in the first year of life. Previous studies have explored the developmental changes in the P300 component of event-related potentials (ERP) measured by electroencephalography (EEG). This component is found in a large variety of tasks in adults, and is specifically associated with the detection of targets and novel objects in a visual task, and considered as an index of the speed of information processing (Duncan-Johnson & Donchin, 1982). In infants, the P300 has a slower latency, which accelerates throughout development, especially throughout the first year of life (De Haan et al., 2003; Fuchigami et al., 1995; Kouider et al., 2013; Nelson & Salapatek, 1986; Riggins & Scott, 2020; Van Dinteren et al., 2014). An acceleration of internal brain rhythm such as the alpha or the mu rhythms has been describe from infancy to adulthood (Bender et al., 2023; Elhamiasl et al., 2023). Also the speed of information transmission increases from infancy to adulthood, as evidenced by the decrease of the latency of an evoked response (Van Blooijis et al., 2023). Such acceleration of the information processing is also reflected in visuo-attentional processes such as the attentional blink, an attentional phenomenon reflected by a miss of a target after reported a first target, when the two targets are presented in a short window. Such process was recently found to be more than 6 times longer at 5 months of age compared to adults, and already about 2 to 3 time faster in 8-month-old infants compare to 5-month-old infants (Hochmann & Kouider, 2022). This acceleration of information processing in infancy is probably related to the myelination of the occipital and parietal lobes, that start in the 4th to 6th month of life, and is, already at 11 months of age, at 50% of its final adults' myelination, in the occipital lobe (Deoni et al., 2011; Dubois et al., 2014).

The encoding of visual features that eventually lead to the representation of visual categories, is done in the ventral stream (Grill-Spector, 2003; Grill-Spector et al., 1998; Grill-Spector & Weiner, 2014; Groen et al., 2017; Konkle & Caramazza, 2013; Reddy & Kanwisher, 2006; Ritchie et al., 2021; Wiggett et al., 2009). In this stream, lower-level features are encoded in more posterior regions, but also decay faster in time compared to higher-level features, that are encoded in more anterior regions (Gao et al., 2020; Groen et al., 2017; Kiebel et al., 2008). Previous results evidenced that younger infants' categorization of objects as animate or inanimate (denoted by looking times) is predicted by the patterns of neural responses to objects in the most anterior part of the adult's ventral stream (Spriet et al., 2022), suggesting that infants mostly rely on higher-level visual features for categorizing objects by animacy, features that decay more slowly as compare to lower-level features. When they grow older,

between 6 to 8 months of age, their categorization performance correlates with information in larger portion of the visual ventral stream (Chapter 4), a timeline that is similar to the acceleration of information processing (Hochmann & Kouider, 2022). Thus, we hypothesized that the acceleration of visual information influences the representation of categories throughout development, and the infants' perceptual awareness of the categories. According to this hypothesis, because younger infants are slower in processing information, and especially here visual information, they would only be able to integrate higher-level visual features when representing objects' category, that decay more slowly as compare to the lower-level visual features, and are encoded in more anterior regions of the ventral stream. Then, as infants get older, and as the speed of information processing accelerates, they would be able to integrate visual features earlier in time increasing the categorization of already represented categories. However, another possibility is that as soon as a category is represented, the speed of information processing does not affect the content of the representation at all. This chapter aims at studying the impact of the acceleration of visual processes on infants' and adults' representation of the visual object categories.

Based on the work in previous chapters, this study focuses on the animate-inanimate categorization. We used the frequency-tagging paradigm (Norcia et al., 2015; Rossion, 2014b) and a regular category-selective stimulation to study the animate-inanimate categorization in adults and infants at 4 and 9 months of age, with different frequencies of images presentation: 4, 6, 12 and 30 Hz (250ms, 167ms, 83ms and 33ms per image respectively). This paradigm is suitable for infants' studies as the response is specified *a priori* by the stimulation frequency of the oddball, easily extracted from the signal noise as not being so much influenced by random (i.e., not regular) artifacts such as movements, and as it can be recorded with short images presentation, allowing to have a response even though the infants' attention is very limited (Kabdebon et al., 2022; Peykarjou, 2022).

We show evidence for an automatic and fast visual categorization of animate and inanimate objects in all age groups, starting from 4 months of age. This categorical response consistently decreased as the stimulation frequency increased for all age groups, but accelerates through development. In particular, it was only observed with the 4 Hz presentation, but not with the 6 Hz presentation in 4-month-old infants; it was found at both 6 and 12 Hz in 9-month-old infants, and in 6, 12 and 30 Hz in adults. We thus measured a dramatic acceleration of the categorization processes, that is at least 3 times faster at 9 months of age compare to 4 months of age, and get even faster later on, until adulthood. While younger infants need time to process objects and learn about them, older infants

get faster in retrieving information based on the visual features into a better understanding of the environment, up until adulthood.

5.2. Materials and Methods

5.2.1. Participants

The study involved a total of 128 infants and 36 adult participants. Experiment 1 involved 36 healthy adults (3 groups of 12 adults; 15 identified their gender as male, 21 as female, mean age 27.7 ± 6.7 years) with normal or corrected-to-normal vision and no report of history of psychiatric or neurological conditions. An additional participant was tested but excluded from the analysis for falling asleep during the experiment. The sample size for adults' experiment ($N = 12$) was chosen based on a power analysis of the results of experiment with similar design (see Chapter 6) that indicates a minimum of 3 participants with this set of stimuli (Cohen's $d = 3.826$, $\beta = 0.95$, $\alpha = 0.05$, one-tail t -test; GPower 3.1). Experiment 2 involved 64 infants of 4 months divided in 2 groups of 32 infants each (35 male, 29 female; mean age 138.3 ± 9.3 days), and 10 additional infants were tested but excluded from the analysis for crying and moving too much, leading to not enough trial included (see preprocessing). Experiment 3 involved 64 infants of 9 months divided in 2 groups of 32 infants each (32 male, 30 female; mean age 288.0 ± 8.9 days) with 7 additional infants being tested but excluded from the analysis for not looking at the screen, moving too much or pulling off the net during the experiment, leading again to not enough trial included. The sample size for infants' experiments ($N = 32$) was chosen based on a statistical power analysis performed on prior work using frequency-tagging to study the categorization of human faces (de Heering & Rossion, 2015). In that study, 15 infants of 4 months of age showed evidence of categorizing face stimuli. The minimal sample size for the observed effect was estimated to be 9 infants (Cohen's $d = 1.298$, $\beta = 0.95$, $\alpha = 0.05$; GPower 3.1). In the present study, we expected a smaller effect size as we were studying categories with higher within-category variability than the human face category. Moreover, we aimed at comparing 2 speed conditions at each age. Thus, we decided to test 32 infants per group. After having tested the first group, we verified that this sample size was larger than the minimal sample size ($N = 19$) required to obtain the effect (group of 4-month-olds tested with 4 Hz presentation: Cohen's $d = 0.610$, $\beta = 0.95$, $\alpha = 0.05$; GPower 3.1). and keep the same sample size.

5.2.2. Stimuli

Stimuli were colored photographs of isolated real-world objects picked from the internet, with the intention to reflect as much as possible the actual variability of each category in the real world. All stimuli were superimposed on a gray background, and each image was resized at 629x629 pixels. Images depicted 320 different animate (51 fish, 70 birds, 14 amphibians, 6 turtles and 179 non-human mammals including 10 nonhuman primates, 5 cats and 8 dogs) and 320 different inanimate objects (16 constructions, 108 graspable objects such as tools, 54 fruits and vegetables, 43 flowers bushes and trees, 66 pieces of furniture and 33 vehicles). Both animate and inanimate stimuli contained different views of the objects, i.e., objects could be in front-view, but also seen from the top, or the side. Not all images were centered into the square, again aiming at increasing the variability of the stimulus set. For a better description of the stimuli, see Chapter 6, “original set”.

5.2.3. Experimental procedure

Adult participants sat on a chair ~60 cm away from a 60 Hz computer screen (resolution 1920x1200 pixels, size 51.5x32.2cm), where stimuli were presented centrally (16° of visual angle). Infant participants sat on their parents’ lap, in the same apparatus adults were. Stimuli were presented via a fast periodic visual stimulation (FPVS), at the frequency of 4, 6, 12 or 30 Hz (250, 166.67, 83.33 or 33.33 ms per image), using Psychtoolbox (Brainard, 1997) and MATLAB (The MathWorks, Nantucket, MA) in a squarewave design.

Experiment 1. Each participant was tested in one condition only (6, 12 or 30 Hz) with 32 trials. Each trial lasted 32s and consisted of one stream of images, pseudo-randomly selected from one category (animate or inanimate) with the regular insertion, every five images, of one image of the other category, leading a category-selective stimulation of $6/5 = 1.2$ Hz; $12/5 = 2.4$ Hz or $30/5 = 6$ Hz (Figure 1 A). Half of the trials used animate as the rare, category-selective stimulation, and half used inanimate. Each trial started by a fade-in (increase in contrast) of 2s and ended by a fade-out (decrease in contrast) of 2s, to ease the presentation for participant eyes and avoid eye movement caused by an abrupt appearance or disappearance (de Heering & Rossion, 2015). At the end of each trial, participants were presented with one image from the regular category (e.g., animate) to avoid creating an attentional bias toward the rare category (e.g., inanimate), as attention could enhance categorization (Kaiser, Oosterhof, et al., 2016). Participants were asked to say if they saw the image in the previous stream, by pressing one of two keys. This task was only designed to help participant to stay focus on the screen and the experiment, and results were not analyzed.

Experiments 2-3. The procedure for infants' group was identical to that of Experiment 1, except for the following aspects. First, 4-month-olds were tested with 4 Hz or 6 Hz streams (Experiment 2) and 9-month-olds with 6 or 12 Hz streams (Experiment 3). For the 4 Hz streams, the rare category images occurred every 4 images instead of every 5 images, leading a category-selective frequency of $4/4 = 1$ hz (Figure 1 A). Second, infants only saw a maximum of 16 trials, which all shared the same rare category (either animate or inanimate). The experiment ended when all the 16 trials were shown, or when the infant started to cry, fell asleep, moved a lot or stopped looking at the screen anymore. Finally, infants did not take the memory test.

5.2.4. EEG

EEG recordings. EEG data was acquired using 128-channel EGI nets (Electrical Geodesics, Inc.). Data were acquired with vertex reference, using the EGI Net Station acquisition software, continuously digitized at a sampling rate of 1kHz (net amp 400 system EGI). Impedance was lowered for each participant as much as possible, not exceeding 40k Ω . During the experiment, triggers were sent from the experimental computer to the acquisition computer via a light sensor: a white square appeared in correspondence to the sensor (i.e., at the bottom right of the screen) at the beginning and at the end of each trial.

EEG preprocessing. All preprocessing steps were performed using EEGLAB toolbox (Delorme & Makeig, 2004) and MATLAB R2015b. Raw data for each participant were first filtered using a 4th-order high pass butterworth filter at 0.1 Hz and a 4th-order low pass butterworth filter at 100 Hz. When needed, electrodes were interpolated (no more than 3 electrodes per adult participant and 1 electrode per infant participant). For adult participants, all electrodes were re-referenced using the average of all electrodes as a reference. For infants, we removed outer band electrodes due to the presence of increased noise (i.e. electrodes 1, 8, 14, 17, 21, 25, 32, 38, 43, 48, 49, 56, 63, 68, 73, 81, 88, 94, 99, 107, 113, 119, 120, 121, 125, 126, 127, and 128), mostly due to movement artifacts. The remaining 100 electrodes were re-referenced using their average as a reference. Finally, recordings were segmented by trials, taking 25s of each trials (25, 30, 60 or 150 complete cycles for respectively 4, 6, 12 and 30 Hz presentation) starting 2.5s after the beginning of the trial, except for the 4 Hz presentation for which the divided trials started 3s after the beginning of the actual trial.

For infants' data, we only considered trials for which we were certain infants paid attention to the sequence of images. For this, we applied a Fast Fourier Transform (FFT) to each trial, over the 3 occipital electrodes, where we expect to find high visual response to the image streams. We averaged

the spectra of those 3 electrodes, per trial and per participant, and computed the signal-to-noise ratio (SNR) by dividing the amplitude at the frequency of the image presentation (i.e., 4, 6 or 12 Hz) and 2nd harmonic (i.e., 8, 12 or 24 Hz), by the local average signal, that is the mean of amplitudes of the 20 (for 4 Hz presentation rate) or 24 (for 6 and 12 Hz presentation rate) surrounding bins (10 or 12 bins on each side of the amplitude, excluding the immediate adjacent bins). If the SNR at the 1st and/or 2nd harmonic of the image presentation frequency was superior to an arbitrary threshold of 2, as previously used in infants EEG experiment of frequency tagging (de Heering & Rossion, 2015; Peykarjou, 2022), the trial was included in further analyses.

In addition, we excluded trials for which the experimenter noted online that the infant moved too much, did not look at the screen, touched the mother with their head, pulled off, touched or grabbed the net, or cried. Infants that provided only one trial were not included in the final sample (in total, eleven 4-month-old infants and eight 9-month-old infants). On average, 4-month-old infants tested with 4 Hz streams contributed 6.75 ± 3.53 trials, 4-month-old infants tested with 6 Hz streams contributed 6.75 ± 4.28 trials, 9-month-old infants tested with 6 Hz streams contributed 6.41 ± 3.64 trials and 9-month-old infants tested with 12 Hz streams contributed 6.19 ± 2.80 trials.

Included trials were averaged in the time domain for each participant, each electrode and each trial type (animate rare category and inanimate rare category). We then applied a FFT to those averages for examination in the EEG frequency-domain at the high frequency resolution of 0.04 Hz ($=1/25$ seconds). EEG spectra were averaged across trial types (for adult participants), leading to one EEG spectrum per participant, at each electrode. The baseline-corrected amplitudes were computed for each participant and electrodes by subtracting to the amplitude of interest the local baseline, that is the mean of amplitudes of the surrounding bins (10 bins on each side for the 4 Hz presentation; 12 bins on each side for other presentations). That corresponds to a frequency range of ± 0.8 and ± 0.96 Hz. The number of bins were reduced for the 4 Hz presentation to avoid that the same bin was used in computing the baseline-corrected amplitude of 2 different harmonics of the category-selective frequency. The baseline-corrected amplitudes were also computed for all bins surrounding each frequency of interest, and used to visualize the data (Figure 2).

5.2.5. Analysis

Group analyses. To select the harmonics and electrodes to further include in the analyses, grand-averaged spectrum (average EEG spectrum of the group) was computed for each adult groups by averaging together all the participants' spectrum in the time domain for each type of trial and electrode,

applying the FFT, averaging the spectra across the type of trials and finally averaging the spectra across all electrodes. For infants' analyses, we only considered the 27 posterior electrodes. On this grand-averaged spectra, z-scores were computed as the difference between the amplitude at the frequency of interest and the average amplitude of the surrounding bins divided by the standard deviation of the surrounding bins excluding the maximum and minimum (Quek & Rossion, 2017; Retter & Rossion, 2016). For each category-selective frequency (i.e., 1, 1.2, 2.4 or 6 Hz), all harmonics below 50 Hz with a z-score higher than 1.64 (one-tailed, $\alpha = .05$) were selected. When analyzing the category-selective response, harmonics that are common to the visual presentation rate (i.e., 4, 6, 12 or 30 Hz and harmonics) were not considered). After having identified harmonics that may show an effect, electrodes were selected based on the z-score computed at each electrode. For each electrode, z-scores at all previously identified harmonics were averaged, and this average was compared to a threshold at 3.33 (one-tailed, $\alpha = .0004$, Bonferroni correction for 128 electrodes) for adults or 2.88 (one-tailed, $\alpha = .0019$, Bonferroni correction for 27 posterior electrodes) for infants.

To assess whether the amplitude at the frequency of interest and harmonics was significantly higher than the noise level for each group, a one-tailed one-sample *t*-test tested the sum of baseline-corrected amplitudes at the identified harmonics averaged over the identified electrodes for each participant (Retter et al., 2021), against 0 (noise level). In addition, we evaluated whether each participant, individually, exhibited a category-specific response, defined as z-scores above 1.64 for the first harmonic (1, 1.2, 2.4 or 6 Hz depending on the condition) on at least 2 neighboring electrodes.

Comparisons between groups. The responses measured in the different groups were also compared in each experiment, performing a cluster-mass permutation analysis. We looked for differences between groups computing the sum of baseline-corrected amplitudes over all harmonics of the category-selective frequency under 50 Hz for each electrode and participants. For each electrode, we performed a one-way ANOVA to compare the three adult groups (Experiment 1), or two-sample *t*-test to compare infant groups (Experiment 2-3). Neighboring electrodes showing an above threshold effect ($P < .05$) were clustered and tested for significance using nonparametric cluster-mass permutation test (Maris & Oostenveld, 2007) with 5000 random permutations of the condition labels of the original data.

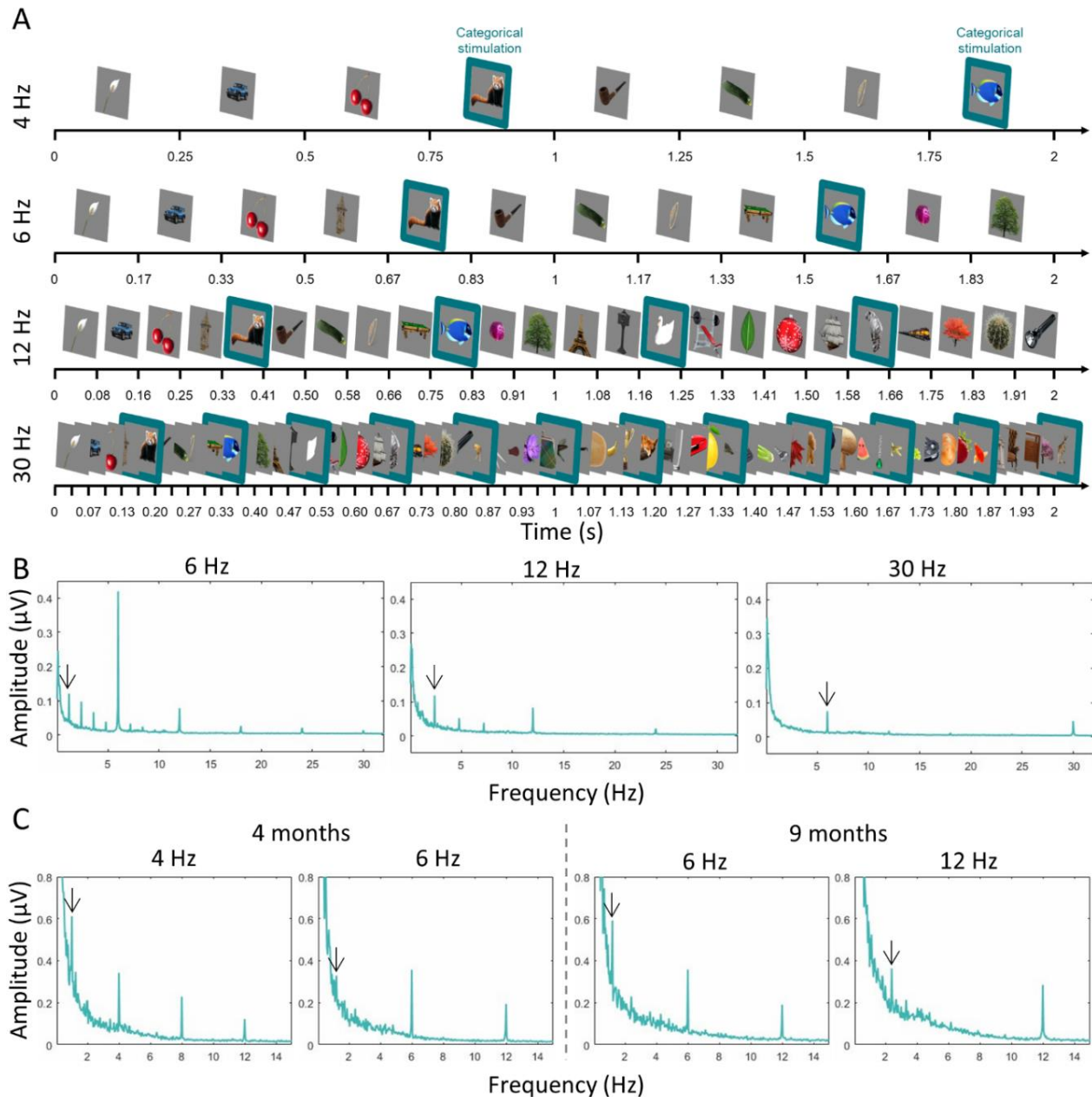


Figure 1. Experimental design and frequency spectrum. **(A)** Two seconds extracts from streams of visual stimuli testing for the categorization of animals among inanimate objects at different speed of presentation: 4, 6, 12 and 30 Hz. Images squared in blue are the animate categorical stimulation. **(B)** Frequency spectrum of adults, denoting visual synchronization with the different speed of presentation, as seen by peaks at 6 Hz (left panel), 12 Hz (middle panel) and 30 Hz (right panel), as well as visual categorization of animals and inanimate objects at Ft (pointed by arrows). Frequency spectra are computed over the average of all electrodes. **(C)** Infants' frequency spectrum for 4-month-olds (left) and 9-month-olds (right), denoting visual synchronization with the different speed of presentation, as seen by peaks at 4 Hz (left panel – 4-month-olds), 6 Hz (right panel – 4-month-olds; left panel – 9-month-olds) and 12 Hz (right panel – 9-month-olds), as well as visual categorization of animals and inanimate objects at Ft (arrows). Frequency spectra are computed over the average of posterior electrodes.

5.3. Results

5.3.1. Experiment 1

Experiment 1 tested three groups of adult participants ($N = 12$ per group) with streams presented at 6, 12 or 30 Hz, and one image from the rare category every five images (Figure 1). Each group showed responses largely distributed over the scalp and peaking over posterior electrodes (6 Hz group: $M_{Amplitude} \pm sd = 0.284 \pm 0.160$; 95% CI = 0.201 – Inf; $t(11) = 6.149$; $P < .0001$; $d = 1.775$; 12 Hz group: $M_{Amplitude} \pm sd = 0.211 \pm 0.085$; 95% CI = 0.167 – Inf; $t(11) = 8.623$; $P < .0001$; $d = 2.489$; 30 Hz group: $M_{Amplitude} \pm sd = 0.108 \pm 0.085$; 95% CI = 0.064 – Inf; $t(11) = 4.405$; $P < .001$; $d = 1.272$; Figure 2 A). All participants exhibited an individual significant response at the category-selective frequency, as shown by the individual baseline-corrected amplitudes (Figure 2 A).

A cluster-mass permutation test relying on a one-way ANOVA compared the baseline-corrected amplitudes for different conditions (i.e., different group), summed for all harmonics of the category-selective frequency under 50 Hz, at each electrode. This analysis revealed a significant cluster of 86 electrodes where the category-selective responses varied depending on the speed of stimuli presentation ($P < .0001$). In that cluster of electrodes, the category-selective response did not differ for 6 Hz and 12 Hz streams ($M_{Difference} \pm sd = 0.087 \pm 0.137$; 95% CI = -0.029 – 0.203; $t(22) = 1.562$; $P = .133$; $d = 0.638$), but it was higher for both of these conditions than for 30 Hz streams (6 Hz vs. 30 Hz streams: $M_{Difference} \pm sd = 0.215 \pm 0.131$; 95% CI = 0.104 – 0.326; $t(22) = 4.007$; $P < .001$; $d = 1.636$; 12 Hz vs. 30 Hz streams: $M_{Difference} \pm sd = 0.128 \pm 0.076$; 95% CI = 0.063 – 0.193; $t(22) = 4.101$; $P < .001$; $d = 1.674$).

We thus observe a decrease in the category-selective response resulting from an increase in the speed of the stream of images from 12 to 30 Hz. This pattern reflects weaker integration of visual features into visual categories at high presentation speeds.

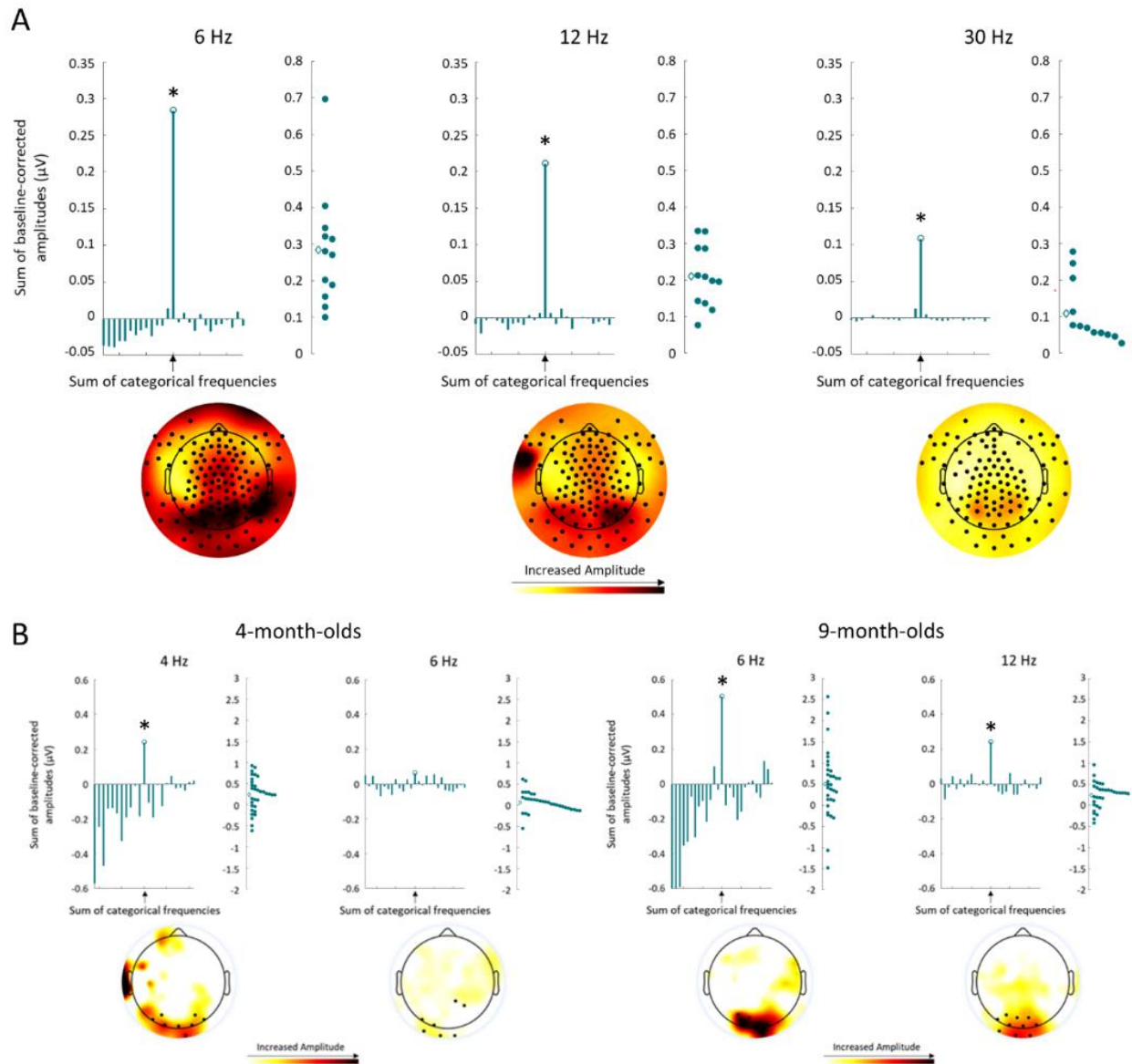


Figure 2. Category-selective response for stimulus streams of different speeds and at different ages. Results of Experiment 1 (A) and Experiments 2-3 (B). Sum of baseline-corrected amplitudes for the selected harmonics, averaged over the selected electrodes and all participants of each group. Baseline-corrected amplitude responses were computed subtracting the average amplitude of surrounding frequencies to the amplitude of the category-selective response. On the right of each sub-panel, each dot indicates the response of one participant, and the diamond indicates again the group-level average. Selected electrodes for these analyses are indicated by dots on the topoplots below each graph. *: Different from chance $P < .05$.

5.3.2. Experiment 2

Experiment 2 and 3 examined how slow processing of visual information in infancy (Hochmann & Kouider, 2022) impacts the capacity of infants to categorize visual stimuli as animate or inanimate

objects, as a proxy of their ability to make sense of the visual input. Experiment 2 tested one group of 4-month-old infants ($N = 32$) with stream of images at a rate of 4 Hz and one image of the rare category every 4 images (category-selective frequency: $4/4 = 1$ Hz), and a second group of 4-month-olds ($N = 32$) with streams of images at a rate of 6 Hz and one image of the rare category every 5 images (category-selective frequency: $6/5 = 1.2$ Hz). We observed a significant category-selective response over 8 posterior electrodes (Figure 2 B) for the 4 Hz streams ($M_{Amplitude \pm sd} = 0.240 \pm 0.393$; 95% CI = $0.122 - \text{Inf}$; $t(31) = 3.449$; $P < .001$; $d = 0.610$). No significant category-selective response was observed for the 6 Hz streams ($M_{Amplitude \pm sd} = 0.063 \pm 0.230$; 95% CI = $-0.006 - \text{Inf}$; $t(31) = 1.547$; $P = .066$; $d = 0.273$). At the individual level, 11 infants (34%) showed a significant category-selective response with the 6 Hz streams, whereas 21 infants (66%) showed a significant category-selective response with the 4 Hz streams. These distributions differed significantly (Fisher's exact test CI = $1.299 - 10.226$; OR = 3.645 ; $P = .024$).

A cluster-mass permutation test relying on a t -test comparing the baseline-corrected amplitudes of the spectrum summed for all the harmonics of the category-selective frequency below 50 Hz at each electrode revealed a significant cluster of 7 posterior electrodes where the category-selective response for the 4 Hz streams was higher than the category-selective response for the 6 Hz streams ($P < .0001$). In sum, 4-month-olds' categorization capacities decreased strongly when streams accelerated from 4 Hz to 6 Hz.

5.3.3. Experiment 3

Experiment 3 tested one group of 9-month-old infants ($N = 32$) with streams of images at a rate of 6 Hz and one image of the rare category every 5 images (category-selective frequency: $6/5 = 1.2$ Hz), and one group of 9-month-olds ($N = 32$) with streams of images at a rate of 12 Hz and one image of the rare category every 5 images (category-selective frequency: $12/5 = 2.4$ Hz). We observed a significant category-selective response over 5 posterior electrodes (Figure 2 B) for the 6 Hz streams ($M_{Amplitude \pm sd} = 0.503 \pm 0.805$; 95% CI = $0.261 - \text{Inf}$; $t(31) = 3.533$; $P < .001$; $d = 0.625$), and a significant category-selective response over 10 posterior electrodes (Figure 2 B) for the 12 Hz streams ($M_{Amplitude \pm sd} = 0.242 \pm 0.301$; 95% CI = $0.152 - \text{Inf}$; $t(31) = 4.550$; $P < .0001$; $d = 0.804$). At the individual level, 19 infants (59%) showed a category-selective response with the 6 Hz streams and 24 (75 %) with the 12 Hz streams. These distributions did not differ significantly (Fisher's exact test CI = $0.168 - 1.416$; OR = 0.487 ; $P = .287$).

A cluster-mass permutation test relying on a *t*-test compared the baseline-corrected amplitudes of the spectrum summed for all harmonics under 50 Hz at each electrode for the two speed conditions. This analysis revealed no cluster of electrodes with different category-selective response for the 6 Hz and 12 Hz streams, suggesting that an acceleration of the visual streams from 6 to 12 Hz had little impact on 9-month-old infants' categorization capacities.

5.3.4. Comparison at different stages of development

We compared the responses of 4- and 9-month-olds to 6 Hz streams in Experiments 2 and 3. A cluster-mass permutation test relying on two-sample *t*-test at each electrode compared the baseline-corrected amplitudes summed over all harmonics of the category-selective frequencies under 50 Hz and identified a cluster of 4 posterior electrodes showing a larger category-selective response at 9 months than at 4 months of age ($P < .001$; 4 months; $M_{Amplitude \pm sd} = 0.018 \pm 0.524$; 9 months; $M_{Amplitude \pm sd} = 0.514 \pm 0.742$).

5.4. Discussion

Overall, adult participants effectively showed a categorization response for images presented at 6, 12 or 30 Hz, though the categorization response decreased significantly at 30 Hz (Experiment 1). Four-month-olds showed categorization response for images presented at 4 Hz, but not for images presented at 6 Hz (Experiment 2), whereas 9-month-olds showed categorization response for images presented at 6 Hz or even 12 Hz. These results show that infants are capable of integrating visual information in order to extract the fundamental visual categories corresponding to animate and inanimate objects, even at a very young age, despite the large variability of those categories and the limited direct experience of our young urban participants with animals. Additionally, we have demonstrated that as infants grow older, from 4 to 9 months of age, and later, to adulthood, their categorization response get faster and faster (Figure 2 B). At 4 months, infants appeared about 7 times slower (4 Hz vs. 30 Hz) than adults. By 9 months, infant perception had greatly accelerated by a factor of about 3; the category-specific response was qualitatively and quantitatively equivalent at 4 Hz for 4-month-olds and at 12 Hz for 9-month-olds. The observed acceleration along development parallels the estimated acceleration of visuo-attentional processes causing event segmentation and perceptual sampling (factor of 6 between 5 months and adults; factor of 2 to 3 between 5 and 8 months (Hochmann & Kouider, 2022)). This

congruence suggests that the acceleration of attention is shaping the efficiency and rapidity of visual categorization.

5.4.1. The animate-inanimate categorization

Categorization is a fast process, that occurs as soon as human adults detect objects in a visual scene (Grill-Spector & Kanwisher, 2005). The animate-inanimate categorization is the first to be represented in infancy (Spriet et al., 2022) and perhaps the most important distinction for the survival of many species, as it allows a rapid detection of a potential predator or prey in the environment. Animacy acts as an organizing principles for both grammar and discourse (Dahl, 2008), influence human memory (Bonin et al., 2014; Hagen & Laeng, 2017; Nairne et al., 2013, 2017), and visuo-attentional capacities (Guerrero & Calvillo, 2016; Lindh et al., 2019). Earlier, children also interact more with live animals than toys (LoBue et al., 2013), and infants look more at animate compare to inanimate stimuli (Spriet et al., 2022). Infants also show different expectations for animals and inanimate objects. For instance, they are surprised when animates, but not inanimate objects are empty inside (Kominsky et al., 2021), and they do not expect an object to move by itself (Kuhlmeier et al., 2004; Saxe et al., 2005). They also understand that agents are more likely to have goals and intentions (A. Woodward, 1998). All those studies evidence the importance of the animate-inanimate categorization, in shaping adults, children and infants' interaction with the world. Our results suggest that this specific animate-inanimate categorization is already fast and automatic in young infants, as soon as 4 months of age, and accelerates throughout development, based on the infants' speed of integration of the visual features.

5.4.2. The consequences of slower categorization abilities

What could be the consequences of slow vision in young infants? To get a sense of what it means for infants to see slowly, one might consider the interaction between the actual speed of scene in the real-world, and the rate of image capture of a camera that would captured the scene. If both speeds are identical, we perceive events at their actual pace, and looking at the actual scene or at its recording wouldn't change anything. But if the capture is slower than the actual scene, we shall perceive accelerated events, faster changes. We would have difficulties to understand everything the same way when looking at the recording as compare to the actual scene. Analogously, slow perception in early infancy might lead to the experience of a fast-moving, fast-changing world. Thus, if infants' perception is seven times slower than adults' perception, infants would experience a world that moves and changes seven times *faster* than the world adults experience. Consequently, infant perception may be unable to

process certain fast moving events, which may be filtered out, sparing more stable, enduring properties of the world. Slow perception in infancy may paradoxically constitute an efficient feature of early development, reminiscent of the “less is more” hypothesis that highlighted the potential role of limited executive functions in early learning (Newport, 1990). By starting slow, perception highlights the basic and stable structure of the world at the beginning of life. As perception accelerates with development, it gradually introduces more detailed and variable events. The acceleration also happens in time for the development of locomotion (Adolph & Hoch, 2019; Campos et al., 2000). Before they crawl or cruise, infants’ visual input consists of relatively long and stable episodes. People and objects pass through their visual field, but the frame and landscape vary little, unless an external force (typically an adult) moves the baby. As infants begin to crawl, typically around 9 months (DeMasi et al., 2023), they self-generate a more variable visual input, segmented in shorter episodes and requiring faster perception.

It follows from these observations that a deregulation of the typical acceleration in the course of development could have major implications. One recent study suggested faster perceptual sampling in autistic children relative to neurotypical peers (Freschl et al., 2021). Could fast perception be a causal factor of autistic disorders? Autistic children appear to integrate visual information within shorter time windows than their non-autistic peers. This would lead to overly detailed visual representations, as highlighted in recent theories of autism (e.g., Happé & Frith, 2006; Lawson et al., 2015; Mottron & Burack, 2001; Pellicano & Burr, 2012; Van de Cruys et al., 2014). In sum, typical symptoms of autism could result from temporal alterations in the information process, which in turn alter the amount of information intake, with catastrophic cascading effects. Studying infants speed of categorization might be a direction for futur studies in autism and their speed of perception.

5.4.3. Conclusion

To conclude, the present work opens a breach in the mystery that constitutes young infants’ experience of the world, constraining future theories of learning, typical and atypical cognitive development. What is it like to be a baby? How does it feel to sense the world through an infant’s body, to interpret it with an infant’s brain? How much of their input can infants exploit for learning? These problems appear as intractable as understanding what it is like to be a member of another species, such as a bat or a fish (Nagel, 1974). But human babies being human, their experience of the world may not be that foreign from our own human adult experience. Infant perception may be conceived as an altered, not yet mature, version of adult perception. We showed that infants’ vision is much slower than adult vision, leading to the experience of a faster-moving and faster-changing visual world. As perception

accelerates in the course of development, the perceived world slows down. Infants' perception can thus be simulated by combining appropriate filters reflecting variations in visual acuity and color perception with an acceleration of the video frame rate. These simulations provide a unique glimpse into infants' experience of the world and novel insights on the birth of human learning and human thinking.

6. Animate-inanimate object categorization from minimal visual information in the human brain and deep neural networks

This chapter is focusing on human adults' categorization ability to categorize objects into the large, superordinate categories of animate and inanimate. The aim is to study which features are necessary and sufficient to see animacy. We use the same stimuli set as in the previous study, and transformation of this set, so that images were unrecognizable and only retain few, low- and mid-level features, encoded at different level in the adults' ventral stream, survive the transformation. We further explore infants' categorization capacity with stimuli retaining low-level features as well; these results can be appreciated in the supplementary results, in Chapter 9, after the bibliography.

6.1. Introduction

An astonishing capacity for *life detection* underlies animals' survival. Humans and other animals are endowed with mechanisms for rapid and accurate discrimination between animate and inanimate *things*. Accordingly, categorization by animacy is a main organizing principle of object-related information in mind/brain (Caramazza & Shelton, 1998; Grill-Spector & Weiner, 2014; Martin, 2007; Warrington & Shallice, 1984), the first to emerge in infancy (Spriet et al., 2022), and one of the most efficient mechanisms of visual perception (New et al., 2007; Thorpe et al., 1996; VanRullen & Thorpe, 2001).

Categorization by animacy is resolved, for the large part, on the basis of visual information. While motion is a critical cue to distinguish animate from inanimate objects, even for static stimuli, this distinction is computed within milliseconds (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Proklova et al., 2019). Yet, the features that are necessary and sufficient to *see* animacy remain debated (Bracci & Op De Beeck, 2023; Jozwik et al., 2022; Thorat et al., 2019).

Object categorization is the result of tuning to complex visual features in higher-level visual areas, but new results show that low- or mid-level features also carry information about animacy (Grootswagers, Robinson, Shatek, et al., 2019; Jagadeesh & Livingstone, 2024; Kramer et al., 2023; Li & Bonner, 2020; Lieber et al., 2024; Long et al., 2017; Wang et al., 2022). Among the mid-level features which are encoded in middle areas of the visual ventral stream and would capture distinctive properties of animacy are shape and texture (Schmidt et al., 2017; Tiedemann et al., 2022). The role of

these features had, for instance, been highlighted using *texforms* stimuli that only retain the texture and global form of real-world objects, and yet elicit effects similar to those associated with the animate-inanimate distinction in visual cortex (Kramer et al., 2023; Long et al., 2017, 2018; Wang et al., 2022). Lower-level features such as color, encoded in posterior areas of the ventral stream, can also carry information about animacy (Rosenthal et al., 2018).

Many features contribute to the animate-inanimate distinction, but what exactly accounts for the fast, automatic categorization that supports life detection? We used frequency-tagging electroencephalography (ftEEG), to capture the fast and automatic neural response locked to a stimulus appearance. We tested whether such response to visual object perception already distinguishes between animate and inanimate and, if so, which features – low-, mid- or high-level – are sufficient to observe such categorization-response.

In ftEEG, stimuli presented in rapid sequence at a regular frequency, elicit steady-state visual evoked potentials (SSVP) at the same frequency, which allegedly capture the immediate, automatic response to stimulus perception (Liu-Shuang et al., 2014; Norcia et al., 2015; Rossion, 2014b; Rossion & Boremanse, 2011). We presented stimuli at 6 Hz with a regular categorical change at 1.2 Hz, such that every five exemplars of a category (e.g., inanimate), an exemplar of another category (i.e., animate) was presented. We expected a peak of activity at 6 Hz; furthermore, if animates were readily represented as distinct from inanimate objects, we should observe a distinctive signal at 1.2 Hz, corresponding to the categorical change. This approach has been used to show automatic, relatively narrow, distinctions such as faces *vs.* non-face stimuli (e.g., Rekow, Baudouin, Durand, et al., 2022; Rossion et al., 2015), or categories of natural *vs.* artificial objects (Stothart et al., 2017). Here, we considered the broad distinction between animate and inanimate (and the largest stimulus set so far), testing whether, say, a zebra, a fish, a hamster and a turtle, however visually different, are readily *seen* as more similar to each other than to a hammer, a rock, a flower and a plane, and *vice versa*. Moreover, in different conditions and experiments, we manipulated the original images to test whether mid-level (global form, texture) or low-level features (spectral power, color, contrast, luminance, number of pixels) alone could elicit fast and automatic categorization by animacy. The same question was studied using behavioral judgments obtained from human participants, and the categorization performance of VGG-19 (Simonyan & Zisserman, 2015) and GoogLeNet (Szegedy et al., 2015). In these two deep convolutional neural networks (DNNs), which provide successful artificial models for human visual object recognition, we tested how different sets of images were classified in different layers.

6.2. Experiment 1

6.2.1. Materials and Methods

Participants.

Experiment 1 involved twelve healthy adults (7 identified their gender as female, 5 as male, mean age 25.7 ± 4.7 years). An additional participant was tested and excluded from the final sample for falling asleep during the experiment. Without prior data, Experiment 1 was exploratory with respect to the sample size. Results were used for a power analysis to select the sample size for the next experiments. Participants had a normal or corrected-to-normal vision and reported no history of psychiatric or neurological conditions. This and the following experiments were approved by the local ethics committee (CPP Ile de France VIII).

Stimuli.

Experiment 1 involved two sets of stimuli: 640 intact colorful images (hereafter, *original set*), and the corresponding phase-scramble version.

Original set. Stimuli were created from 640 colorful photographs of real-world animate ($n=320$) and inanimate ($n=320$) objects taken from the internet, representing as much as possible the diversity of objects in the real world. Each object was presented in a box of 629×629 pixels with a gray background (Figure 1A). The animate set included 51 fish, 70 birds, 179 nonhuman mammals, 14 amphibians and 6 turtles, all different from one another. Humans were excluded to prevent a bias in the animate-inanimate categorization. Insects, spiders and reptiles (except for turtles) were also excluded to prevent emotional reactions (e.g., disgust or fear). The inanimate set included 223 artificial and 97 natural objects. Among the natural objects, there were 54 different fruits and vegetables, and 43 different flowers, bushes and trees. Among the artificial objects, there were 16 different exemplars of buildings and constructions, 108 different exemplars of clothes, pieces of jewelry, buttons, coins, and tools, 66 pieces of furniture and 33 vehicles. Objects could appear in all sorts of orientations and visual angles. From these images, two types of stimulation sequences were created (animate and inanimate), named after the category that served as *oddball*, i.e., the less frequent stimulus-category used to elicit the periodic categorization response (Figure 1B). Thus, in the animate-sequence, 320 inanimate objects were shown as *standard* (i.e., the frequent stimulus-category in the sequence), and 68 animate objects

were shown as oddballs; in the inanimate-sequence, 320 animate objects were shown as standard, and 68 inanimate objects were shown as oddballs.

Phase-scramble stimuli. Phase-scramble stimuli were created manipulating the spatial spectrum of the images in the original colorful set through phase-scrambling, using the *imscramble* function (<http://martin-hebart.de/webpages/code/stimuli.html>) run in Matlab (The Mathworks, Natick, MA). This manipulation preserved the number of pixels, color, contrast, luminance and power-spectral distribution of the original images, but removed mid-level (shape and texture) information effectively rendering the object identity unrecognizable (see below).

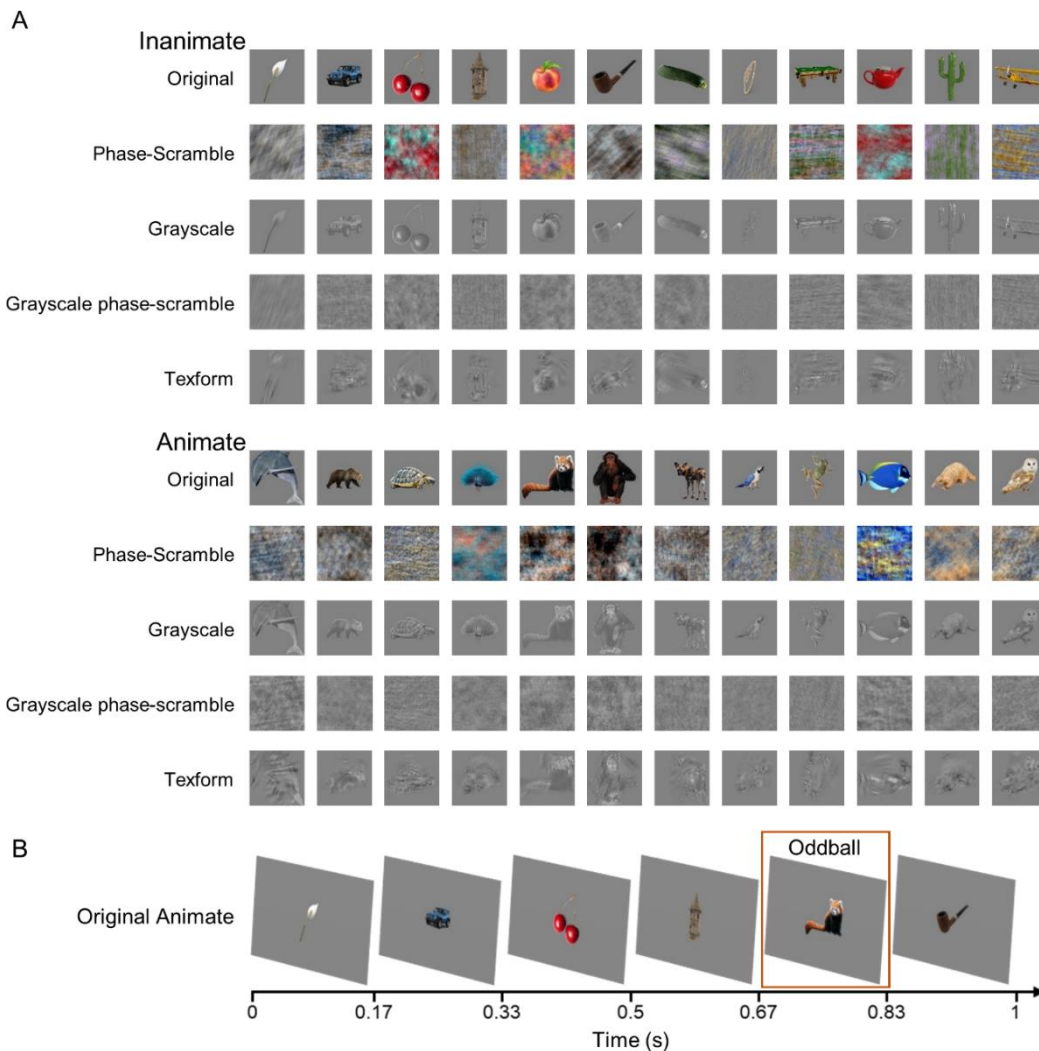


Figure 1. Illustration of stimuli and experimental design in Experiments 1-3. (A) Examples of stimuli from the *original set* (Experiment 1) in the inanimate condition (top) and animate condition (bottom) and the different versions of those images: phase-scramble (Experiment 1), grayscale (Experiments 2-3), grayscale phase-scramble (Experiment 2) and texforms (Experiment 3). (B) One-second extract from the “animate” stimulation sequence based on the *original set*, testing for the categorization of animates (*oddball* category) among inanimate objects (*standard* category).

Procedures.

Participants sat on a chair ~60 cm away from a 60 Hz computer screen (resolution 1920x1200 pixels, size 51.5x32.2cm), where stimuli were presented centrally (16° of visual angle). Stimulus presentation was controlled using Psychtoolbox (Brainard, 1997) through Matlab. Each participant was tested in two conditions involving the original and the phase-scramble set, respectively. Each condition included 16 trials of animate-sequences (i.e., animate stimuli used as oddball among inanimate stimuli), and 16 trials of inanimate-sequences (i.e., inanimate stimuli used as oddball among animate stimuli). Therefore, each participant completed a total of 64 trials (16 trial x 2 conditions x 2 types of sequences). Each trial started by a fade-in phase (increase in contrast) of 2 s and ended with a fade-out phase (decrease in contrast) of 2 s, to ease the stimulus presentation for the participant's eyes and avoid eye movements caused by abrupt appearance or disappearance. Each trial lasted 32 s, and consisted of a sequence of stimuli, presented at a base frequency of 6 Hz (6 images per s; 166.67 ms per image), in a squarewave design, where every 5 images of one category (standard), one image of the other category (oddball) was presented. With this trial structure, a categorization response was expected at 6/5, i.e., 1.2 Hz (Figure 1). The five standard stimuli that preceded each oddball were pseudo-randomly selected to prevent the repetition of the same image in a trial and the presentation of two images from the subordinate-level category at the same frequency as the oddball presentation (e.g., a dolphin and a cat, both from the mammal category, presented with only five images in between). For a participant, the same list of images was used in the “original” and “phase-scramble” conditions.

The 16 trials of a condition were presented in a single block, yielding a total of four blocks (animate and inanimate sequences in the original and phase-scramble conditions). The two blocks belonging to the same condition were presented one after the other, with the order of type of sequence (animate or inanimate) and condition (original or phase-scramble) counterbalanced across participants. At the end of the trial, a test-image from the standard category was shown. Participants were instructed to pay attention to each image in a trial and to report whether the final test-image was shown during the last trial, by pressing “A” or “P” (“yes” or “no”, counterbalanced between participants). The test-image was chosen from the standard category to prevent an attentional bias toward the oddball category. This task was included to encourage attention towards the stimuli. The whole experiment lasted ~45 minutes.

EEG.

EEG recording. EEG data were acquired using 128-channel EGI nets (Electrical Geodesics, Inc.). Data were acquired with vertex reference, using the EGI Net Station acquisition software, continuously

digitized at a sampling rate of 1kHz (net amp 400 system EGI). Impedance was lowered for each participant as much as possible, not exceeding 40k Ω . During the experiment, triggers were sent from the experimental computer to the acquisition computer via a light sensor: a white square appeared in correspondence to the sensor (i.e., at the bottom right of the screen) at the beginning and at the end of each trial.

EEG preprocessing. All preprocessing steps were performed using the EEGLAB toolbox (Delorme & Makeig, 2004) and MATLAB R2015b. Raw data for each participant were filtered using a 4th-order high pass butterworth filter at 0.1 Hz and a 4th-order low pass butterworth filter at 100 Hz. All electrodes were re-referenced using the average of all electrodes as a reference. Data were segmented by trials, taking 25 s for each trial (30 complete cycles) starting 2.5 s after the trial onset. Trials were averaged for each participant, for each condition (original and phase-scramble) and for each type of sequence (animate and inanimate). A Fast Fourier Transform (FFT) was applied for data examination in the EEG frequency-domain at the high frequency resolution of 0.04 Hz (1/25 seconds). Baseline-subtracted amplitudes were computed for each participant, for each condition, at each electrode, and for each harmonic of the response (base and oddball) by subtracting from the amplitude of interest, the average amplitude of the local baseline, that is, the mean of amplitudes of the 24 surrounding bins (12 on each side excluding the immediate adjacent bins for a frequency range of ± 0.48 Hz) excluding the maximum and minimum (Retter and Rossion, 2016; Quek and Rossion, 2017). To define the peaks at the frequencies of interest (harmonics of the base and oddball responses), we averaged the EEG signal of all participants in the time domain to obtain grand-averaged spectra for each condition, at each electrode. We then applied the FFT, and averaged the sequence type. We then computed z-scores for each condition, first on the average over all electrodes, by subtracting from the amplitude of interest, the average amplitude of the local baseline and dividing this difference by the standard deviation of the local baseline. Electrodes showing an effect were identified computing the z-score at each electrode and for each condition.

EEG data analyses.

Data were analyzed considering the two image-type conditions (original and phase-scramble), averaging across the sequence type (animate, inanimate), after applying the FFT (see preprocessing).

First, we identified the harmonics showing a response as follows. We computed the z-score (see above) for all harmonics of the oddball frequency under 12 Hz (excluding the base frequency at 6 and 12 Hz), and for all harmonics of the base frequency under 50 Hz. We selected harmonics with z-

score higher than a threshold of 1.64 ($\alpha = .05$, one-tailed). Second, we identified the electrodes at which a significant response was found. Again, we computed a z-score on the grand-averaged spectrum but this time for each electrode independently, for the harmonics previously selected. We averaged those z-scores and tested them against a threshold of 3.33 (corresponding to $\alpha = .0004$, one-tailed, Bonferroni correction for 128 electrodes).

We used parametric statistical tests to address the following effects. First, we assessed whether the amplitude of the response in each condition separately (base response or oddball response) was significantly higher than the noise level. To this end, we used a one-tailed t-test testing the baseline-subtracted amplitude, summed over the identified harmonics, averaged over the identified electrodes for each participant (Retter et al., 2021), against 0 (noise-level). Second, we compared the two conditions (original vs. phase-scrambled image type) using a paired t-test testing the difference between baseline-subtracted amplitudes, summed over the identified harmonics, averaged over the identified electrodes, for each participant and each condition. Finally, we ran a 2 Image Type x 2 Sequence Type repeated-measures ANOVA to test whether the oddball response in the conditions with the original and the phase-scramble set was affected by the type of sequence (animate or inanimate).

Data availability.

Stimuli, EEG data and code for the main analyses will be deposited in the Open Science Framework repository created for this project.

6.2.2. Results

Experiment 1 tested whether the fast presentation of a sequence of images from a large and heterogeneous set, could elicit a response in correspondence to a regular categorical change from animate to inanimate object, and *vice versa*. Furthermore, by using phase-scramble images, we tested whether such categorization-response could be elicited by stimuli that only preserved low-level visual features of animate and inanimate objects (*i.e.*, power-spectrum, color, contrast and luminance).

Base-frequency response. We first verified that we could measure a response at the base stimulation frequency of 6 Hz in each condition. All harmonics below 50 Hz showed a significant response ($z > 1.64$). The response was widely distributed over the scalp (Figure 2A) for both the original condition ($M_{Amplitude \pm sd} = 0.802 \pm 0.360$; 95% CI = 0.615 – Inf; $t(11) = 7.714$; $P < .0001$; $d = 2.227$) and the phase-scramble condition ($M_{Amplitude \pm sd} = 0.771 \pm 0.348$; 95% CI = 0.591 – Inf; $t(11) = 7.671$; $P < .0001$; $d = 2.215$). This result showed reliable synchronization of electrical activity with the visual stimulation at 6 Hz, in both conditions.

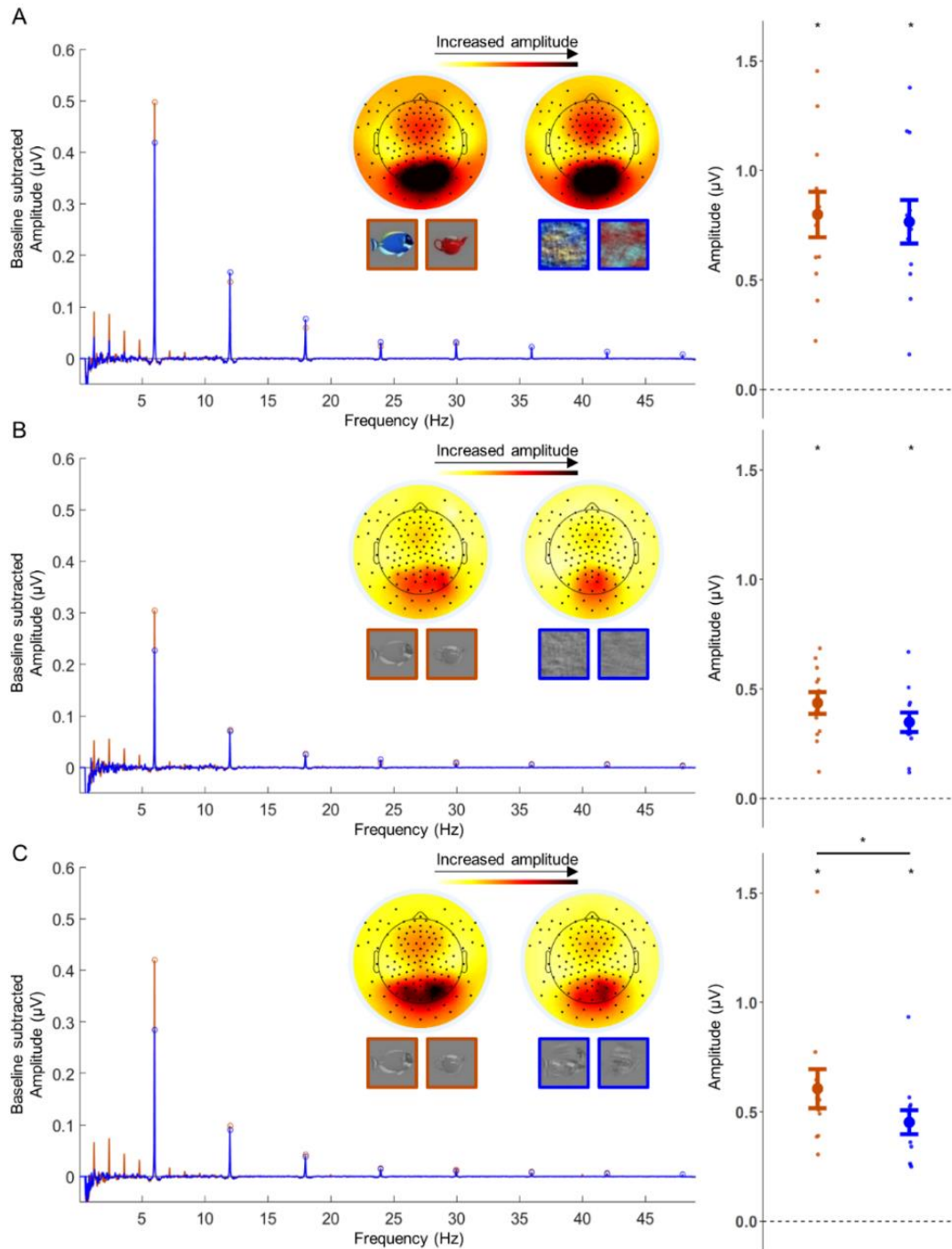


Figure 2. Response at the base frequency. (A) **Experiment 1.** *Left:* Baseline subtracted amplitudes as a function of frequencies in the conditions with the original set (in orange) and the phase-scramble set (in blue). Circles signal the significant harmonics of the base response, at the group level. The scalp distribution of baseline subtracted amplitudes summed over harmonics show a widespread response, peaking over posterior electrodes. *Right:* Baseline subtracted amplitudes summed over harmonics and averaged over electrodes for each condition (original in orange and phase-scramble in blue). Small dots represent participants' individual response, thicker dots represent the average responses, lines represent the standard error of the mean. (B) **Experiment 2.** *Left:* Same measures as in Fig. 2A left but for the conditions with grayscale stimuli (orange) and grayscale phase-scramble stimuli (blue). *Right:* Same measures as in Fig 2A right but for the conditions with grayscale stimuli (orange) and grayscale phase-scramble (blue). (C) **Experiment 3.** *Left:* Same measures as in Fig. 2A left but for the conditions with grayscale stimuli (orange) and texforms (blue). *Right:* Same measures as in Fig. 2A but for the conditions with grayscale stimuli (orange) and texforms (blue).

Oddball response. We analyzed the response at the oddball frequency, which we took as a signature of automatic visual categorization by animacy. This response was first compared against the noise-level separately for each Image-Type condition (original and phase-scramble), and then compared between the two conditions. For the original condition, we found nine harmonics showing a significant widespread categorization-response (Table 1), peaking over posterior electrodes ($M_{Amplitude} \pm sd = 0.319 \pm 0.083$; 95% CI = 0.276 – Inf; $t(11) = 13.254$; $P < .0001$; $d = 3.826$; Figure 3 A). For the phase-scramble condition, we found a significant categorization response in the first four harmonics, spreading over the scalp ($M_{Amplitude} \pm sd = 0.126 \pm 0.059$; 95% CI = 0.096 – Inf; $t(11) = 7.480$; $P < .0001$; $d = 2.159$). This response was further evidence in an additional group of 4-month-old infants (see Supplementary Results). A t -test revealed a difference between conditions, with greater categorization-response for the original-set condition ($M_{Difference} \pm sd = 0.205 \pm 0.100$; 95% CI = 0.142 – 0.269; $t(11) = 7.108$; $P < .0001$; $d = 2.052$). This effect was further replicated in 4-month-old infants (see Supplementary Results).

Finally, the 2 Image Type x 2 Sequence Type ANOVA showed that the difference in the categorization-response between the two Image type conditions was not affected by the type of sequence; that is, a significant categorization was found irrespective of whether the sequence involved animate oddballs among inanimate standards, or *vice versa*. In fact, only the effects of Image Type was significant, $F(1,11) = 51.006$, $P < .0001$, but there was no effect of Type of sequence, $F(1,11) = 1.146$, $P = .307$, or interaction, $F(1,11) = 0.959$, $P = .349$.

Table 1. Z-scores of the 8 first harmonics, excluding the 5th, for the category-selective frequency and for each condition and sub-condition of Exp. 1, Exp. 2 and Exp. 3

Exp.	Condition	1.2 Hz	2.4 Hz	3.6 Hz	4.8 Hz	7.2 Hz	8.4 Hz	9.6 Hz	10.8 Hz
1	Original	6.170	13.400	23.622	13.717	9.372	11.232	2.853	3.432
	Phase-scramble	1.877	6.943	6.824	4.062	0.886	0.854	-0.146	-2.009
2	Grayscale	3.148	6.205	7.002	7.598	12.047	7.254	4.189	2.903
	Grayscale Phase-scramble	-0.231	1.538	2.884	0.881	-3.189	0.818	-0.060	-0.284
3	Grayscale	4.635	24.176	17.457	16.135	11.941	6.316	3.702	2.915
	Texform	0.721	2.454	0.939	0.960	3.536	1.233	-1.050	0.654

Note: Exp., experiment; Highlighted in bold are the significant results (z -scores > 1.64).

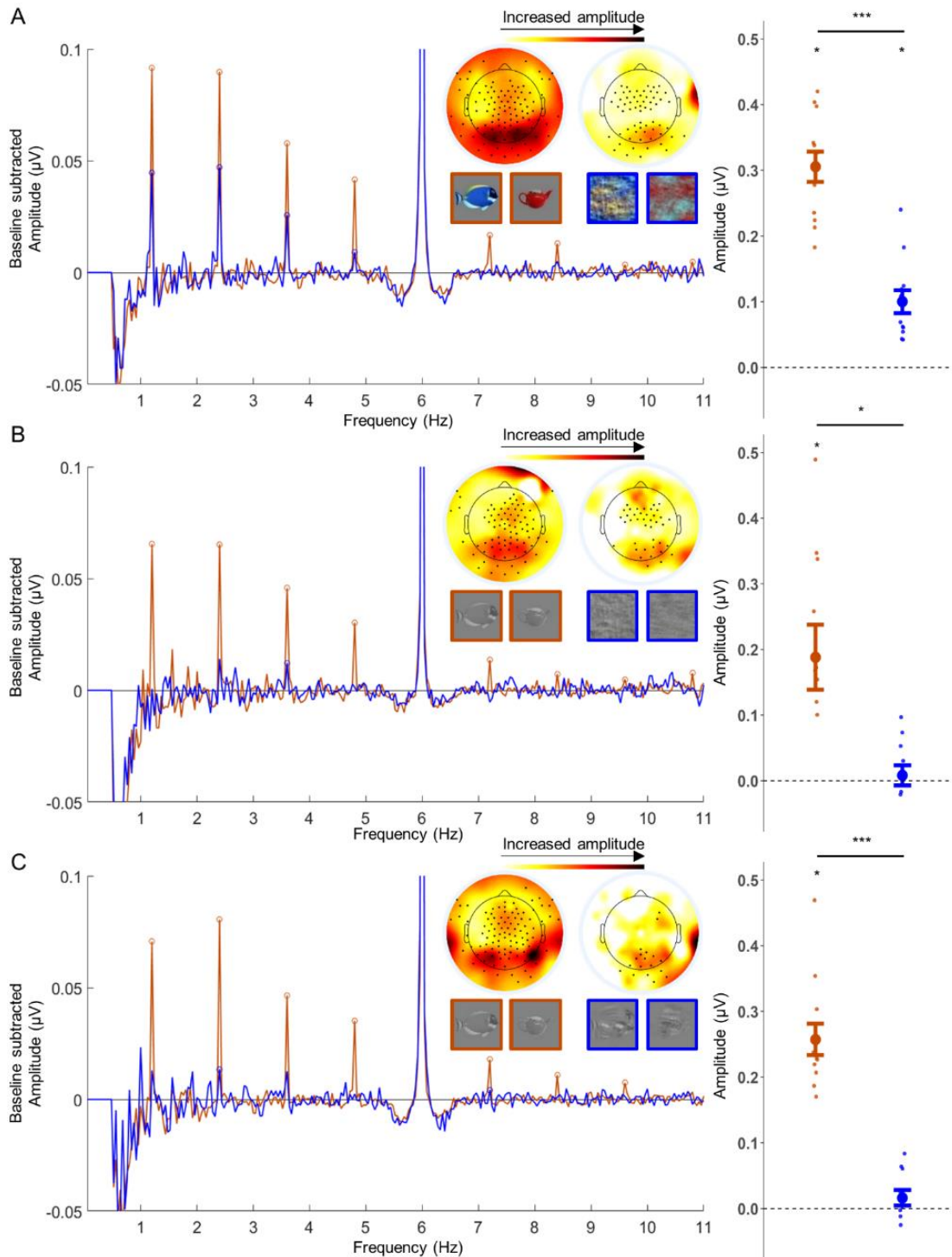


Figure 3. Response at the oddball frequency. (A) Experiment 1. *Left.* Baseline subtracted amplitudes as a function of frequencies below 12 Hz, for the conditions with original stimuli (orange) and phase-scramble stimuli (blue). Circles signal the harmonics with significant oddball response. The distribution over the scalp of the summed harmonics baseline subtracted amplitudes show a widespread response peaking over posterior electrodes. *Right.* Baseline-subtracted amplitudes summed over harmonics and averages over electrodes show a difference between conditions. **(B) Experiment 2.** Left and right panels show the same measures as Fig. 3A, but for the conditions with grayscale stimuli (orange) and phase-scramble grayscale stimuli (blue). **(C) Experiment 3.** Left and right panels showed the same measures as Fig. 3A, but for conditions with grayscale stimuli (orange) and texforms (blue).

6.3. Experiment 2

6.3.1. Materials and methods

Participants.

Experiment 2 involved twelve healthy adults (9 identified their gender as female, 3 as male, mean age 26.3 ± 6.3). One additional participant was tested and excluded from the final sample for falling asleep during the experiment. A power analysis based on the results of Experiment 1 estimated that a sample size of 7 was required to obtain the smallest categorization effect observed in Experiment 1 (phase-scramble condition; Cohen's $d = 1.421$, $\beta = 0.95$, $\alpha = 0.05$; GPower 3.1). Therefore, we decided to keep the same sample size of 12 across all the experiments. Participants had normal or corrected-to-normal vision and reported no history of psychiatric or neurological conditions.

Stimuli, procedures and analyses.

In Experiment 2 everything was identical to Experiment 1 except for the stimuli: the original set in Experiment 1 was modified using the `step1_lumContrastOriginals_fromGreen` function (<https://github.com/brialorelle/texformgen>), to create grayscale versions of each image (hereafter, grayscale set). From these grayscale versions, the corresponding (grayscale) phase-scrambled versions (hereafter, grayscale phase-scramble set) were created following the procedure described in Experiment 1, with animate and inanimate stimuli matched for luminance and contrast (Figure 1A). In addition, we ensure grayscale phase-scramble stimuli were unrecognizable (e.g., a fish that could not be recognized as “fish”, a flower that could not be recognized as “flower” and so on). Recognition was evaluated using the data of a naming task administered to an independent group of native-French speaker participants ($n=19$). For this task, participants were informed that they would see transformed versions of existing objects, and had to name each object, using one word, or more if needed. Stimuli were shown in the center of the screen (16° of visual angle – assuming a distance of ~ 60 cm from screen) until the participant responded. Underneath each image, there was a box in which the participant could type the response. Images were presented in a random order. An image was considered to be unrecognizable when no more than one participant named it correctly at the basic-level. Only 2 grayscale phase-scramble images were recognizable, as being correctly named by 2 participants – a bird and a stone. Thus, we considered the phase-scramble sets as not recognizable.

6.3.2. Results

Results of Experiment 2 replicated those of Experiment 1, with animate and inanimate stimuli that lacked color information and were matched for contrast and luminance.

Base-frequency response. Results revealed a significant response over all harmonics, widely distributed over the scalp (Figure 2 B) for both conditions (grayscale condition: $M_{Amplitude} \pm sd = 0.439 \pm 0.171$; 95% CI = 0.351 – Inf; $t(11) = 8.927$; $P < .0001$; $d = 2.577$; grayscale phase-scramble condition: $M_{Amplitude} \pm sd = 0.359 \pm 0.156$; 95% CI = 0.278 – Inf; $t(11) = 7.983$; $P < .0001$; $d = 2.305$).

Oddball-frequency response. Response at the oddball-frequency was first compared against noise-level separately for each Image Type condition (grayscale and grayscale phase-scramble) and then compared between the two conditions. For the grayscale condition, nine harmonics showed significant categorization response ($M_{Amplitude} \pm sd = 0.240 \pm 0.153$; 95% CI = 0.161 – Inf; $t(11) = 5.424$; $P < .001$; $d = 1.566$; Table 1; Figure 3 B). For the phase-scramble grayscale condition, a significant categorization response was found in the third harmonic ($M_{Amplitude} \pm sd = 0.012 \pm 0.017$; 95% CI = 0.003 – Inf; $t(11) = 2.459$; $P = .016$; $d = 0.710$; Table 1; Figure 3 B). A *t*-test revealed a significant difference, with greater categorization-response in the grayscale than in the grayscale phase-scramble condition ($M_{Difference} \pm sd = 0.180 \pm 0.171$; 95% CI = 0.071 – 0.289; $t(11) = 3.640$; $P = .004$; $d = 1.051$).

Finally, the 2 Image Type x 2 Sequence Type ANOVA only revealed an effect of Image Type ($F(1,11) = 13.066$, $P = .004$) but no effect of Sequence Type ($F(1,11) = 0.013$, $P = .913$) or interaction ($F(1,11) = 0.232$, $P = .640$).

6.4. Experiment 3

6.4.1. Materials and Methods

Participants.

Experiment 3 involved twelve healthy adults (9 identified their gender as female, 3 as male, mean age 23.7 ± 2.7 years). An additional participant was tested and excluded from the final sample for falling asleep during the experiment. Participant had normal or corrected-to-normal vision and reported no history of psychiatric or neurological conditions.

Stimuli, procedures and analyses.

In Experiment 3, everything was identical to Experiment 1, except for the stimuli. Here, stimuli consisted of the same grayscale images used in Experiment 2, and corresponding *texforms* (Figure 1A). Six-hundred-forty *texforms* were created from the 640 grayscale images following the method in Deza et al. (2019). We then selected the inanimate *texforms* ($n=175$) and animate *texforms* ($n=195$) that were unrecognizable. As for grayscale phase-scramble stimuli in Experiment 2, *texform* stimuli recognition

was evaluate evaluated using the data of a naming task administered to another independent group of native-French speaker participants (n=19). We chose to include only texforms that were not recognizable as our goal was to measure categorization that relied on the visual features only (without/before semantic categorization). Based on this selection, we obtained 195 animate and 175 inanimate texforms. To match the number of items between conditions, we randomly removed 20 animate images. The corresponding grayscale images used to create the texforms were also included in Experiment 3. A subset of 47 images in each category were used as oddball stimuli.

6.4.2. Results

Experiment 3 tested whether the mid-level visual features that are preserved in the texforms (texture and global form) could elicit an automatic animate-inanimate categorization-response as measured with ftEEG. Results of Experiment 3 replicated those of Experiments 1-2.

Base-frequency response. A significant response at the base-rate frequency was widely distributed over the scalp (Figure 2 B), in both conditions (grayscale: $M_{Amplitude} \pm sd = 0.605 \pm 0.311$; 95% CI = 0.444 – Inf; $t(11) = 6.745$; $P < .0001$; $d = 1.947$; texform: $M_{Amplitude} \pm sd = 0.456 \pm 0.191$; 95% CI = 0.357 – Inf; $t(11) = 8.262$; $P < .0001$; $d = 2.385$).

Oddball-frequency response. The response at the oddball-frequency was first compared against noise-level separately for each Image Type condition (grayscale and texform), and then compared between conditions. For the grayscale condition, nine harmonics showed a significant and widespread categorization-response ($M_{Amplitude} \pm sd = 0.270 \pm 0.087$; 95% CI = 0.224 – Inf; $t(11) = 10.697$; $P < .0001$; $d = 3.088$; Table 1; Figure 3 C). For the texform condition, two harmonics showed a significant response which was the strongest in posterior electrodes ($M_{Amplitude} \pm sd = 0.018 \pm 0.023$; 95% CI = 0.006 – Inf; $t(11) = 2.648$; $P = .011$; $d = 0.764$; Figure 3 C). A t -test revealed difference between conditions, with greater response to grayscale images than texforms ($M_{Difference} \pm sd = 0.241 \pm 0.101$; 95% CI = 0.177 – 0.305; $t(11) = 8.292$; $P < .0001$; $d = 2.394$).

A 2 Image Type x 2 Sequence Type ANOVA only showed an effect of Image Type ($F(1,11) = 70.443$, $P < .0001$), but no effect of Sequence Type ($F(1,11) = 0.027$, $P = .872$), or interaction ($F(1,11) = 3.808$, $P = .077$).

In summary, Experiment 3 replicated the effect of categorization by animacy for grayscale images. A significant, although weaker, categorization-response was found for the texforms that preserved mid-level visual features, while making the objects unrecognizable.

Comparison between Experiments. Experiments 1-3 were compared considering the EEG categorization-responses to different image-type conditions, averaged across the types of sequence (animate, inanimate). For this analysis we included all the harmonics showing a significant categorization-response in at least one condition, summed their baseline-subtracted amplitudes, averaged the 128 electrodes for each participant in each experiment and condition, and performed an ANOVA with Image Type (intact: colorful or grayscale vs. impoverished: phase-scrambled or texform) as a within-subjects factor and Experiment as a between-subjects factor.

The ANOVA revealed the effects of Image Type, $F(1,33) = 96.203$, $P < .0001$, and Experiment, $F(2,33) = 7.566$, $P = .002$, but no interaction between the two, $F(2,33) = 0.775$, $P = .469$. In all three experiments, the categorization-response was larger for the intact condition than for the impoverished condition. Moreover, there was a significant difference in the amplitude of the categorization-response between Experiment 1 and 2 ($M_{Difference \pm sd} = 0.105 \pm 0.075$; 95% CI = 0.041 – 0.169; $t(22) = 3.399$; $P = .003$; $d = 1.388$) and between Experiment 1 and 3 ($M_{Difference \pm sd} = 0.065 \pm 0.047$; 95% CI = 0.025 – 0.105; $t(22) = 3.397$; $P = .003$; $d = 1.387$): Experiment 1 showed the largest categorization-response, which instead did not differ between Experiment 2 and 3 ($M_{Difference \pm sd} = -0.040 \pm 0.073$; 95% CI = -0.023 – 0.102; $t(22) = -1.317$; $P = .202$; $d = 0.538$).

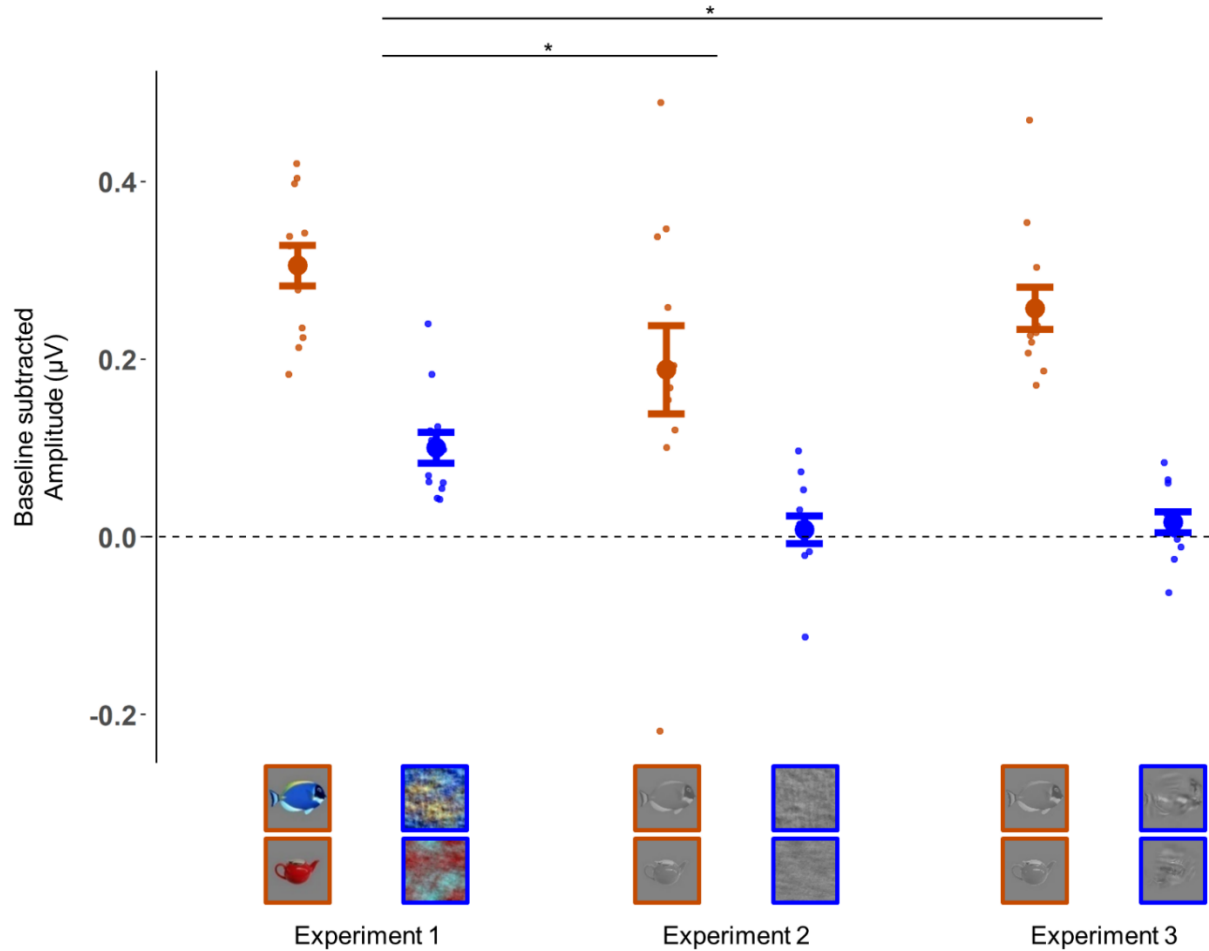


Figure 4. Comparison between Experiments. Sum of baseline subtracted amplitude averaged over electrodes and participants, and standard error of the oddball response harmonics in Experiment 1 (left), 2 (middle) and 3 (right). In orange is the oddball response to intact images (colorful or grayscale) and in blue is the oddball response to impoverished images (phase-scramble or texform). Dots represent individual participants' mean. Horizontal bars and * denote significant effect with $P < .05$.

6.5. Behavioral judgments

6.5.1. Materials and Methods

Participants.

A total of 60 healthy native-English speakers (28 identified their gender as female, 31 as male, 1 as other, mean age 25.4 ± 4.1 years), living in different countries (Africa, America, Asia, Europe) and external to the above EEG study, were recruited online. All participants had normal or corrected-to-normal vision and reported no history of psychiatric or neurological conditions. Each participant performed one of three categorization tasks: 1) animate-inanimate categorization of phase-scramble

stimuli in Experiment 1 (n=20, English speakers, 12 identified their gender as female, 8 as male, mean age 26.6 ± 4 years); 2) animate-inanimate categorization of grayscale phase-scramble stimuli in Experiment 2 (n=20, English speakers, 7 identified their gender as female, 12 as male, 1 as other, mean age 26.5 ± 4.1); 3) animate-inanimate categorization of the 640 texform images used to select stimuli for Experiment 3 (n=20, English speakers, 9 identified their gender as female, 11 as male, mean age 26.7 ± 3.3).

Stimuli and Procedures.

Participants were tested online on Testable.org (Rezlescu et al., 2020). They were asked to sit 60 cm away from the screen (about the length of an arm), align their eyes with the center of the screen, and make an effort to not move during the experiment. They were also asked to turn off their phone and any other device. Prior to the experiment, they gave informed consent and followed instructions for screen calibration. Instructions were as follow: we used a yes-or-no forced-choice task in which participants were instructed to decide whether each image depicted an animal or not. Instructions were in English and, for this task, we recruited English-speaking participants. They were informed that they would see 640 transformed images of existing object, presented one by one and to decide whether the image could be an animal with a yes-or-no response. Stimuli were shown in the center of the screen (16° of visual angle – assuming a distance of ~ 60 cm from screen) until the participant responded. Underneath each image two buttons appeared, for yes or no. The side of the “yes” and “no” buttons were counterbalanced between participants but were always the same for one participant. Participants were instructed to respond by clicking on “yes” or “no” using the mouse. Images were presented in a random order. Three different groups performed the task on the phase-scramble stimuli of Experiment 1, the grayscale phase-scramble stimuli in Experiment 2 and the texforms of Experiment 3, respectively.

Analysis.

Using data from the categorization task, we tested whether even if a specific object in the impoverished set was not recognizable, participants could still *guess* its superordinate-level category (i.e., animate or inanimate) in a forced-choice. To this end, participants’ accuracy rates (i.e., proportions of correct responses) in the categorization task were tested against chance (one-sample *t*-test).

6.5.2. Results

Categorization was at chance for grayscale phase-scramble stimuli ($M_{Accuracy} = 0.503 \pm 0.192$; 95% CI = 0.488 – 0.518; $t(639) = 0.382$; $P = 0.703$; $d = 2.620$), and above chance for colorful phase-scramble stimuli ($M_{Accuracy} = 0.523 \pm 0.187$; 95% CI = 0.508 – 0.537; $t(639) = 3.104$; $P = 0.002$; $d =$

2.797) and texforms ($M_{Accuracy} = 0.628 \pm 0.179$; 95% CI = 0.614 – 0.642; $t(639) = 18.079$; $P < 0.0001$; $d = 3.508$). Moreover, categorization accuracy was significantly higher for texform (texform vs. phase-scramble: $M_{difference} = 0.105 \pm 0.183$; 95% CI = 0.085 – 0.125; $t(1278) = 10.262$; $P < 0.0001$; $d = 0.574$; texform vs. grayscale phase-scramble: $M_{difference} = 0.125 \pm 0.185$; 95% CI = 0.104 – 0.145; $t(1278) = 12.048$; $P < 0.0001$; $d = 0.674$), and did not differ between the two phase-scramble sets ($M_{difference} = 0.020 \pm 0.189$; 95% CI = -0.001 – 0.041; $t(1278) = 1.892$; $P = 0.059$; $d = 0.106$). Results did not change when we considered only the texforms included in Experiment 3 – i.e., above chance categorization ($M_{Accuracy} = 0.592 \pm 0.180$; 95% CI = 0.573 – 0.611; $t(349) = 9.539$; $P < 0.0001$; $d = 3.289$), higher than categorization accuracy for phase-scramble sets (texform vs. phase-scramble: $M_{difference} = 0.073 \pm 0.185$; 95% CI = 0.045 – 0.100; $t(698) = 5.204$; $P < 0.0001$; $d = 0.393$; texform vs. grayscale phase-scramble: $M_{difference} = 0.089 \pm 0.187$; 95% CI = 0.062 – 0.117; $t(698) = 6.319$; $P < 0.0001$; $d = 0.478$). In sum, performance in forced-choice categorization showed that color, contrast, luminance, texture and global form carry information that may aid the recognition of objects as animate or inanimate; among those features, texture and global form, appear to be the most reliable for making that decision.

6.6. DNN data

6.6.1. Materials and methods

DNN.

We selected two convolutional deep neural networks, VGG-19 (Simonyan & Zisserman, 2015) and GoogLeNet (Szegedy et al., 2015), which have been very successful in reaching human-level object recognition, and provide a reliable model of the primate visual system, particularly the ventral temporal cortex (VTC) (e.g., Khaligh-Razavi & Kriegeskorte, 2014; but see Storrs et al., 2020; Kheradpisheh et al., 2016; Kubilius et al., 2016; Yamins et al., 2014). Using these models, we explored whether the different features preserved in the stimulus sets of Experiments 1-3 supported the animate-inanimate categorization at different processing stages, or layers of the visual hierarchy. In particular, we tested whether, depending on the visual features preserved in the different stimulus sets, categorization by animacy emerged in the first, middle or deeper layers of the models, modelling low-, mid- and higher-level visual areas (VTC), respectively.

Multiple processing stages in these DNNs transform input images through a series of nonlinear operations. Initially, convolutional layers apply kernels to small regions of the input image, capturing

spatial hierarchies of visual features. Then, a rectified linear unit (ReLU) function introduces nonlinearity by thresholding activations at zero, allowing only positive activations to pass forward. Then, max pooling layers perform downsampling, which reduces the spatial dimensions of the input while preserving important features. Finally, fully connected layers flatten the processed input into a one-dimensional vector, which represents the class scores. These networks were trained on approximately 1.2 million images from the ImageNet database (ILSVRC2012), encompassing 1,000 classes including animals (40%) and objects (60%). We used pretrained versions of these models available in MATLAB (MatConvNet; Vedaldi and Lenc, 2016), with standardized preprocessing steps such as mean subtraction of the training images and scaling of all stimuli to 224x224 pixels.

Analysis.

For VGG-19, we extracted features from convolutional layers ('conv1_1' through 'conv5_4') and fully connected layers ('fc6', 'fc7', and 'fc8'). For GoogLeNet, we extracted features from initial convolutional layers ('conv1_7x7_s2' and 'conv2_3x3'), various inception modules ('inception_3b_output' through 'inception_5b_output'), and the final classification layer ('loss3_classifier'). For each layer, we computed a dissimilarity matrix (Kriegeskorte, Mur, & Bandettini, 2008) representing the dissimilarity between the features extracted for each of the stimuli used in the EEG study (see Figure 5 for examples of matrices at different layers for VGG-19): original set (Experiment 1), grayscale set (Experiment 2-3), phase-scramble set (Experiment 1), grayscale phase-scramble set (Experiment 2) and texform set (Experiment 3). In a dissimilarity matrix, dissimilarity between each pair of stimuli was computed as $1 - \rho$, where ρ is the *Spearman* coefficient of the correlation between vectors of features extracted for the two stimuli, at a given layer. For each stimulus set, we performed *Pearson* correlations between the matrix extracted from each layer and the reference matrix, corresponding to the dissimilarity matrix extracted for original (intact colorful) stimulus set from the output layer of the DNN ('fc8' for VGG-19 and 'loss3_classifier' for GoogLeNet). This reference matrix was selected as it reflected the performance on the stimulus set with the highest complexity (i.e., highest number of features), and the closest to the stimuli used for training the DNN, thus providing the best representation of the stimuli by the DNN, typically associated with the best performance of the model. Finally, the performance throughout all the layers of a DNN was compared between the different stimulus sets, using pairwise t-tests on the fisher-transformed correlation coefficients.

6.6.2. Results

The purpose of this study was to investigate the representation of the visual features that may distinguish between animate and inanimate objects, across different layers and stimulus sets, using two artificial DNNs. In the EEG study, we reasoned that *intact* stimuli carrying higher-level visual features and *impoverished* stimuli only preserving low- and/or mid-level features (phase-scramble, grayscale phase-scramble and texform), would be processed at different levels of the visual hierarchy – i.e., in higher-level anterior areas *vs.* early or middle areas of the ventral stream. We sought to study whether categorization by animacy could emerge also for impoverished stimuli involving early/middle aspects of the visual stream. Using DNNs, we sought to test the processing differences between stimulus sets across layers and whether and how categorization by animacy is extracted from deeper as well as first and middle layers of the model. Since the results were similar for both models, for simplicity, we only report the results of VGG-19.

First, we tested whether object representation, yielding the broad animate-inanimate distinction as evidenced by the significant correlation of the reference matrix and a model of animacy ($\rho = 0.826$, $P < .001$), could be found when the DNN was tested with the intact as well as the impoverished sets. Providing an affirmative answer to this question, results showed that stimulus information represented by the dissimilarity matrices, increasingly matched the reference matrix as the stimulus complexity increased. However, although correlations were the highest for the intact grayscale set (maximum correlation in layer fc8, $\rho = .817$, $P < .001$), they were also significant for all the impoverished sets (maximum correlation: phase-scramble, layer conv2_1, $\rho = .345$, $P < .001$; grayscale phase-scramble, layer conv2_2, $\rho = .270$, $P < .001$; texform, layer conv2_2, $\rho = .265$, $P < .001$) (Table 2). This means that the categorical animate-inanimate distinction emerged in the DNN also for the most impoverished images. Next, we asked *where* this distinction emerged and peaked for different stimulus sets.

For both intact sets (original and grayscale), the coefficients of the correlation with the reference matrix increased as layers got deeper in the DNN (Table 2), with no difference between layers of the two sets ($M_{\text{difference}} = 0.066 \pm 0.004$; 95% CI = -0.068 – 0.201; $t(36) = 1.001$; $P = 0.324$; $d = 17.223$). This indicates a shift in representation (i.e., a better representation of the images) at the final stages (i.e., deepest layers) of processing. No such increase was observed for the impoverished sets in the last fully connected layers. Here, compared with the intact sets, correlations with the reference matrix across layers were different, reflecting the stability between layers for the impoverished sets as compared to the increase of correlation throughout layers for the intact sets (phase-scramble: $M_{\text{difference}} = 0.157$

± 0.061 ; 95% CI = 0.046 – 0.267; $t(36) = 2.956$; $P = 0.008$; $d = 2.559$; grayscale phase-scramble: $M_{\text{difference}} = 0.207 \pm 0.056$; 95% CI = 0.096 – 0.318; $t(36) = 3.910$; $P = 0.001$; $d = 3.701$; texform: $M_{\text{difference}} = 0.130 \pm 0.047$; 95% CI = 0.020 – 0.239; $t(36) = 2.490$; $P = 0.023$; $d = 2.750$). In effect, for the phase-scramble and texform sets, correlation coefficients did not change much from the first and middle layers to the deeper layers (Figure 5A), implying that the information extracted in deeper layers did not add to the representation of those images in the DNN. In some cases, (i.e., for both phase-scramble sets), correlation coefficients were even higher in middle than in deepest layers, meaning that middle layers were better tuned to the low- and middle-level features carried by those stimuli. Beyond representational differences between different sets of stimuli, these results showed that the animate-inanimate distinction is pervasive in object representation, emerging across different processing stages of the visual hierarchy –not just the deepest layers– and for different, more or less impoverished, visual inputs.

Table 2. Results of the correlations between the reference matrix extracted for the intact colorful set, from the last layer Fc8 (highlighted with the frame), and the matrix extracted for each stimulus set, from each layer of VGG19.

	Original		Grayscale		Phase-scramble		Grayscale phase-scramble		Texform	
Layers	Rho	<i>P</i>	Rho	<i>P</i>	Rho	<i>P</i>	Rho	<i>P</i>	Rho	<i>P</i>
conv1_1	.225	<.001	.136	<.0001	.277	<.001	.190	<.001	.185	<.001
conv1_2	.216	<.001	.131	<.0001	.254	<.001	.183	<.001	.181	<.001
conv2_1	.331	<.001	.183	<.0001	.345	<.001	.238	<.001	.218	<.001
conv2_2	.221	<.001	.242	<.0001	.237	<.001	.270	<.001	.265	<.001
conv3_1	.218	<.001	.220	<.0001	.171	<.001	.136	<.001	.241	<.001
conv3_2	.273	<.001	.243	<.0001	.199	<.001	.126	<.001	.262	<.001
conv3_3	.273	<.001	.244	<.0001	.166	<.001	.107	<.001	.255	<.001
conv3_4	.339	<.001	.257	<.0001	.201	<.001	.109	<.001	.252	<.001
conv4_1	.260	<.001	.224	<.0001	.179	<.001	.088	<.001	.229	<.001
conv4_2	.263	<.001	.237	<.0001	.162	<.001	.107	<.001	.239	<.001
conv4_3	.248	<.001	.232	<.0001	.200	<.001	.143	<.001	.248	<.001
conv4_4	.223	<.001	.201	<.0001	.243	<.001	.216	<.001	.236	<.001
conv5_1	.274	<.001	.227	<.0001	.243	<.001	.189	<.001	.224	<.001
conv5_2	.357	<.001	.283	<.0001	.229	<.001	.176	<.001	.227	<.001
conv5_3	.377	<.001	.327	<.0001	.218	<.001	.168	<.001	.253	<.001
conv5_4	.348	<.001	.331	<.0001	.189	<.001	.171	<.001	.254	<.001
fc6	.745	<.001	.582	<.0001	.187	<.001	.164	<.001	.248	<.001
fc7	.818	<.001	.636	<.0001	.190	<.001	.166	<.001	.261	<.001
fc8	1.000	<.001	.817	<.0001	.145	<.001	.131	<.001	.265	<.001

Highlighted in bold are the significant results; $\alpha = .05$; correlations are with the last layer (Fc8) of the model when classifying the original set of images.

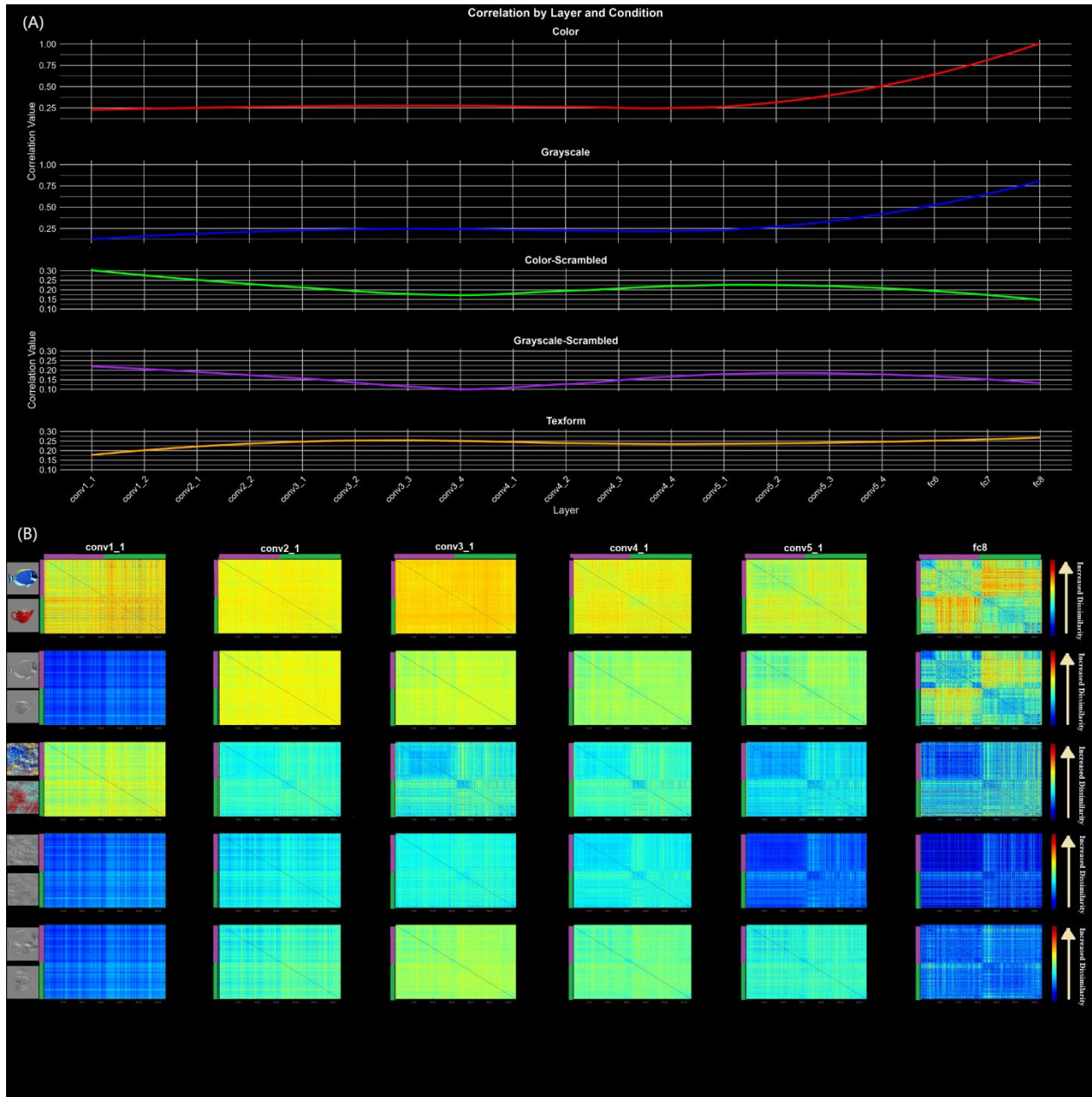


Figure 5. Correlation coefficients and dissimilarity matrices across stimulus sets and layers of VGG-19. (A) Coefficients of the correlation (y-axis) between the representation of a given stimulus set at each layer, and the representation of the intact colorful set in the layer fc8 (x-axis). Each plot corresponds to a stimulus set: intact colorful, intact grayscale, phase-scramble, grayscale phase-scramble, and texform. (B) Examples of dissimilarity matrices extracted from different layers for different stimulus sets. Each row represents a stimulus set, each column corresponds to a different layer. These matrices illustrate representational dissimilarities at different stages of processing. Dissimilarity values are from 0 (lowest dissimilarity, dark blue) to 1 (highest dissimilarity, dark red). This figure illustrates the extent to which the dissimilarity structure changed across layers and conditions.

6.7. Discussion

The animate-inanimate distinction underlies a *life detection* capacity that is crucial for the survival of many animal species. Thus, in an evolutionary perspective, it is not surprising that such distinction emerges right after visual information reaches the visual cortex, suggesting an underlying feedforward mechanism (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Proklova et al., 2019). On the hypothesis that categorization by animacy is primarily a visual process, we used fast periodic visual stimulation combined with ftEEG, to investigate: (1) whether the automatic neural response locked to the stimulus appearance, already carries information about the categorical animate-inanimate distinction, and (2) whether such information can be even extracted from impoverished images, which only retain low- to mid-level features of animate or inanimate objects. To this end, we used an unprecedented set of stimuli, testing the hard task of identifying category membership in a large and heterogeneous set of images, in which animate objects were as varied as mammals, birds, fish and amphibians, and inanimate objects were as varied as furniture, vehicles, tools, plants and vegetables.

ftEEG results demonstrated that information relevant for the animate-inanimate distinction was extracted rapidly from higher-level as well as mid- and low-level visual features. In particular, a strong, widespread response to a regular categorical change in the stream of visual images was found for colorful, *intact*, depictions of real-world objects. This response was only moderately diminished when the same images were in grayscale, and persisted, although weaker, for unrecognizable images that only kept low-level spectral (color, contrast, luminance, number of pixels, power spectrum) information (phase-scramble set) or mid-level (texture and global form) information (texforms). These findings add to a growing body of studies demonstrating that object categorization is the result of tuning to complex visual features in higher-level visual areas, but can be also resolved based on low- and mid-level visual information encoded in early and middle aspects of the visual ventral stream for object recognition (Kramer et al., 2023; Rosenthal et al., 2018; Wang et al., 2022). Furthermore, these findings characterize object recognition as an incremental process in which the accumulation of visual information increases categorization performance, which would be reflected in the amplitude of the categorization response in the EEG signal (see also Carlson et al., 2013; Cichy et al., 2014).

Our analysis of the DNNs confirmed this model. First, we found that intact (colorful or grayscale) images were represented across all layers of the DNN, with increasingly high accuracy, and the highest accuracy in deeper layers, modeling higher-level visual areas (VTC). Instead, for

impoverished (phase-scramble) images, representation was the most accurate in middle layers, suggesting that those layers are better tuned to the low- and middle-level features carried by those stimuli. Second, above and beyond the differences, stimulus representation yielding the animate-inanimate distinction emerged across all layers and for all stimulus sets (the correlations with the “reference matrix” were all significant). This supports the view that the earlier layers of the network, modelling early and middle visual areas, are tuned to distinctions in low- and mid-level features, informative about the animate-inanimate categorization.

Finally, with a behavioral study, we addressed the extent to which features at different levels of complexity could support explicit categorization by animacy in a forced-choice task. The analysis of the participants’ responses showed that, while animate-inanimate classification was at chance for the grayscale phase-scramble set, texforms and colorful phase-scramble images yielded above chance performance (although performance was not reliably different between the two phase-scramble sets). These findings suggest that the categorization response in the EEG signal, as weak as it might be (in the case of phase-scramble and textform sets), is behaviorally relevant, supporting deliberative processing of objects (see also Long et al., 2017). Interestingly, these findings also imply that information about low- and mid-level features gives access to supra-ordinate (animate-inanimate) categories. In the preliminary study of our stimuli, we asked participants to name texforms and phase-scramble images, under the hypothesis that, in naming objects, individuals privilege basic-level category labels (e.g., they name a zucchini “zucchini”, rather than “vegetable” or “inanimate object”; Mervis & Pani, 1980; Mervis & Rosch, 1981; Rogers & Patterson, 2007; e.g., Rosch, 1978; see also Long and Konkle, 2017; Long et al., 2017). Only 2 of the phase-scramble images were named correctly, and only a subset of texforms was named by more than 2 participants (see *Stimuli, procedures and analyses* of Experiment 3). This suggests that, while low- and mid-level features give access to supra-ordinate (animate-inanimate) categories, access to finer-grained categorical distinctions may require more (and higher-level) information. In line with this, recent findings on development of visual object recognition showed that very young (4-month-old) infants can represent the broad animate-inanimate distinction among real-world objects, but the representation of finer-grained (e.g., basic-level) categories develops later based on the progressive recruitment and integration of more and more feature spaces distributed across the visual cortex (Spriet et al., 2022; Chapter 3).

In sum, we asked in this study what accounts for the efficient animate-inanimate categorization that supports life detection. Our results demonstrate that, while higher-level features (e.g., the presence of eyes and limbs) may be necessary for recognizing objects at least at the basic-category level,

information relevant for the animate-inanimate distinction is already extracted from low- and mid-level features, suggesting tuning of early and middle level areas of the visual stream to animacy. Put in another way, the absence of higher-level information in our impoverished sets implies that the animate-inanimate distinction observed for those stimuli did not result from top-down connections, but perception is optimized to classify animate and inanimate stimuli in the early stages of visual processing.

Now, in the light of the present results, *what makes things look animate to humans?* It is clear that a variety of features, from low- to higher-level, participate in this categorization. With our methodological approach, it may also be possible to define a hierarchy of distinctiveness, as roughly captured in Figure 4. If the amplitude of the EEG categorization-response is an index of classification accuracy, then higher-level features carried by the intact sets are clearly more informative than mid- and low-level features, isolated in the impoverished sets. On this reasoning, compared with low-level features in phase-scramble images, mid-level features carried by texforms (global shape and texture) add evidence towards the animate-inanimate distinction (see also Grootswagers, Robinson, & Carlson, 2019; Long et al., 2017, 2018; Wang et al., 2022). Among the low- and mid-level features, color seems to add significant gain in terms of classification accuracy: presence of color increased the categorization-response for the colorful (*vs.* grayscale) intact set, and for the colorful (*vs.* grayscale) phase-scramble set, and yielded above-chance classification in the force-choice task with phase-scramble stimuli (see also Rosenthal et al., 2018). Future research may continue the systematic work to establish the impact of each specific visual feature on categorization. However, it is unlikely that this quest will reveal *one feature* responsible for animacy perception, but that animacy perception would result from the combination of a range of visual features of the kinds investigated here, which can be further disambiguated by co-occurring dynamic and non-visual information.

In conclusion, we showed that the animate-inanimate distinction is pervasive in the processing of visual stimuli, and resilient to the loss of information in the visual input: low- and mid-level features, encoded in early and middle-level aspects of the visual ventral stream for object recognition, are sufficient to elicit the fast, automatic categorical response in the human brain, and can inform deliberation on the broad animate-inanimate distinction. However, impoverished sets induced smaller categorization responses relative to intact objects, and could not be recognized beyond a coarse (forced) animate-inanimate distinction. This contributes to defining visual categorization as an incremental process involving the integration of various features distributed across the visual cortex – or encoded across different layers of a DNN. Based on the present behavioral results and previous research (Spriet

et al., 2022; Chapter 3), this incremental process may explain how finer-grained (e.g., basic-level) categories emerge from broader visual categories such as animate and inanimate.

7. General Discussion

7.1. Summary of findings

In this thesis, I investigated the development of visual categorization, especially in the first year of life. In a series of 4 studies (Chapter 3-6), I asked the following questions: 1) When does the visual categories accounting for the visual system organization in Human adults emerge in infants' behavior? 2) What are the respective role of spontaneous brain maturation and visual experience in the visual saliency of features diagnostic of a category? 3) How does the speed of images' presentation impact the animate-inanimate categorization, throughout life? 4) How do low-, mid- and high-level visual features impact the animate-inanimate categorization in adults?

Chapter 3 revealed that at 4 months of age, infants first explore images based on their physical properties (size, compactness, elongation). This reliance on features that are non-categorical (i.e., features that are not eliciting the representation of a category) overshadowed their representation of objects as animate and inanimate. By 10 months of age, infants' looking behavior was spontaneously guided by the animate and inanimate categories when exploring images, and they lost the guidance by the physical properties of the images. Later on, by 19 months of age, other categories represented in the adults' visual cortex emerged, especially the human-nonhuman categorization, as well as some of the basic-level categories tested (i.e., nonhuman faces and bodies, human bodies and natural small objects). In addition, results showed that while infants' categorization ripened with age, bringing about the representation of more categories, infants' behavior correlated with the activity recorded in ever-larger portion of the adult ventral stream. Altogether, those results suggest that the ability to represent finer-grained categories comes with the ability to integrate more and more visual features, integrated in different portion of the ventral stream in adults. It also demonstrates a switch in relying on physical properties of the images towards being guided by category membership when exploring the environment, between 4 and 10 months of age.

Chapter 4 investigates further the age at which the previous observed switch, from relying on physical properties of the object to relying on category membership, happened. Results showed that this switch happens in two steps. Just like 4-month-olds, 6-month-old infants relied primarily on the physical properties of the images, which overshadowed their representation of animate and inanimate objects. Eight-month-old infants relied on both category membership *and* physical properties of the object. Their behavior also correlated with the activity recorded in the adult ventral stream. Thus,

between 6 and 8 months, the saliency of category membership increases, but the saliency of physical properties (non-categorical features) decreases later, between 8 and 10 months. After having identified the timeline of the increasing saliency of category membership, Chapter 4 explored the role of spontaneous brain maturation and visual experience in this switch. Preterm infants born about 2 months before term and aged 8 months, have a brain maturation equivalent to that of full-term 6-month-olds (their corrected age; i.e., the age they would have if they were born at term), but equivalent amount of visual experience as full-term 8-month-old infants (their chronological age). Thus, comparing 8-month-old preterm infants with 6- and 8-month-old full-term infants informed us about the respective influence of brain maturation and experience in the increased saliency of category membership in infants' behavior. Results revealed that preterm 8-month-olds (i.e., averaged corrected age: 6 months) behaved similarly to full-term 8-month-olds, relying on both the physical properties of the objects and the category membership (animate/inanimate). In addition, preterm's behavior correlated with the organization of the anterior part of the adult ventral stream. This suggests a major role of visual experience in the infants' use of category membership when exploring their environment.

Chapter 5 revealed that the speed of presentation of visual stimuli interacts with the human's brain ability to categorize animate and inanimate objects. Moreover, this categorization process gets faster and faster with age. Relying on frequency-tagging EEG, we showed that the adult brain categorize images presented at 30 Hz, while 4-month-olds require a maximum speed of 4 Hz. The adult brain is thus 7 times faster than the 4-month-old brain. Nine-month-old infants could categorize images presented at 12 Hz, suggesting that the 9-month-old brain is already at least 3 times faster than the 4-month-old brain. Because of a slower perceptual system, we argued that younger infants might experience a world that is changing faster compare to older infants and adults. They might be able to only encode stable structure of the world, i.e., structures that do change slowly. Thus, the speed of infants' categorization impacts the content of infants' representation of the world.

Finally, Chapter 6 studied the role of different visual features in the adult categorization of animate and inanimate objects. Using frequency-tagging EEG, we demonstrated that this superordinate categorization was already extracted from very few, low-level visual features, such as the power spectrum of the images. Yet, the more features in the image, the bigger the categorical response was in adults (and infants', as shown in Supplementary Results) brain. This was further supported by deep neural networks (DNNs) performance with the different impoverished version of the images, showing the best performance with the images containing the greatest amount of features. In addition, lower-level visual features were not sufficient to influence adults' behavior in a forced-choice task. Only when

images contained more mid-level visual features were adults able to recognize their superordinate category, although not their identity. This is suggesting that while lower-level visual features are already diagnostic of animate and inanimate categories, influencing both adults and infants' brain, they are not sufficient to influence behaviors and be used in our everyday life.

7.2. Infants represent visual categories from 4 months of age

Over the years, researchers have used infants' looking behavior as a tool to study infants' cognition. While infants can't speak, they can direct their eyes and explore things surrounding them as they want. As such, infants' look is typically interpreted as revealing where infants' allocate their attention. The longer they look at something mirrors their interest in the stimuli (Aslin, 2007). Categories, as warranting an organization of the world into meaningful units, facilitate our comprehension of every object we could encounter. Moreover, they offer a key mechanism for understanding how humans, and especially developing infants, make sense of the world surrounding them, considering the overwhelming amount of information they have access to. As such, previous research focused on infants' categorization abilities, as a window of infants' cognition, by heavily relying on infants' looking time (e.g., Behl-Chadha, 1996; Mareschal & Quinn, 2001; Pauen, 2002; Quinn, 2004; Quinn et al., 1993, 2001; Quinn & Eimas, 1996, 1998). I also used infants' looking time in this thesis, to study their ability to represent categories. However, the paradigm I used in Chapter 3 and 4 differ from the typical paradigm employed in previous studies. Instead of measuring a difference of looking time between stimuli of 2 different categories after having habituated/familiarized infants to one of the category (Behl-Chadha, 1996; Hunter & Ames, 1988; see for review Mareschal & Quinn, 2001; Rose et al., 1982), I recorded a more spontaneous (i.e., direct, without prior exposition) measure of the difference of looking time between 2 stimuli of same or different categories, precluding a possible online formation of a category during the typical familiarization or habituation phase. The aim was to study what features would spontaneously guide infants' behavior when confronted to new objects; would they spontaneously be guided by categorical features, and which categorical features would mainly guide infants' visual exploration? In addition, this paradigm allowed me to study many different categories (eight categories) in one short experiment, despite infants' short attentional span. Thus, categories observed in Chapter 3 and 4 were more likely reflected categories infants spontaneously represent in their everyday life, where they are facing lots of different objects from lots of different category at once. I further targeted categories that appeared as good candidates to be represented by

young infants, by considering 8 basic-level categories, encompassing 4 superordinate categories as well as the big animate-inanimate categorization, all found as eliciting different brain activation not only in the human adults' ventral stream, but also in other species, such as macaques (Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Op de Beeck et al., 2008; Sigala et al., 2002), which suggests a biologically pre-determined organization, enhancing chances to develop early on in life.

7.2.1. Infants' gaze is guided by category membership

When considered all together, results of Chapters 3, 4, 5 and 6 (Supplementary Results) all revealed that young infants, as soon as 4 months of age, already represent objects as animate and inanimate, despite the huge variability within each category. However, category membership was not yet salient enough to guide infants' exploration in a simple preferential looking task. In particular, Chapters 3 and 4 revealed that young infants, at 4 and 6 months, looked at images based on category membership only when the images were presented at the same size. When image size was not controlled, 4- and 6-month-old infants were looking at images according to their size, elongation and compactness, looking at the bigger, less elongated and most compact images. Thus, this paradigm reveal that although infants, as soon as 4 months of age, already represent the animate and inanimate categories, they do not yet rely on those categories when exploring objects surrounding them. Chapters 3 and 4 further replicated the highly documented preference for human faces, from birth (Farroni et al., 2005; M. H. Johnson et al., 1991; see for review Nelson, 2001; Valenza et al., 1996). Whenever a human face was presented to infants, they would prefer to look at this face, no matter what was presented next to it, reflecting the importance of human face compared to anything else. However, when 2 human faces were presented at the same time, young infants would discriminate them and prefer to look at one of the two.

Infants only started to spontaneously rely on category membership by 8 months of age, and this development was mainly linked to infants' visual experience, as preterm infants aged of 8 months were showing a similar behavior. Older infants, at 10 and 19 months of age, did not exhibit any reliance on the physical properties of objects, but only on category membership, particularly the animate-inanimate categories. Nineteen-month-old infants also represent the human and non-human categories, as well as human bodies, non-human animal bodies, non-human animal faces and natural small objects. In addition, 8- 10- and 19-month-old infants' behavior reflected the organization of large portions of the adults' ventral stream. Results from 8- to 19-month-old infants suggest that once infants have a better

distinction of animate and inanimate categories and represent and rely on more categories than the broad animate-inanimate categorization, they represent more and more features integrated in different portion of the ventral stream.

As visual experience account for the infants' reliance of categorical features when exploring objects (Chapter 4), exploring the changes between 6 and 8 months of age can inform about how categorization and behavior are interconnected. Between 6 and 8 months of age, infants develop in many aspects. In particular, a striking development is their development of motor skills. Indeed, infants start to sit around 6 months (Marcinowski et al., 2019), sometimes creep on their stomach (see for review Malina, 2004), but do not move very far from where their caregiver put them, and rely on their caregivers to be placed in a seating position. They also start to reach and grasp objects, heading for small ones that would be more convenient for them to hold (Libertus et al., 2013). While those motor skills develop, they offer more and more possibilities for infants to explore objects, using their hand, reaching to objects, but also visually, especially when sitting, as it provides a vertical view compare to lying that provides an horizontal view (Libertus & Hauf, 2017; Marcinowski et al., 2019; Soska & Adolph, 2014). It is interesting to see that before those motor skills fully develop, and before infants move in their environment by themselves, they do not rely on category membership, despite being able to represent some of them already (animate and inanimate evidence in 4 and 6 months, Chapters 3-5). Most of 8-month-old infants are able to sit by themselves, start to crawl on their knees and hand, some even pull on tables or couch to stand up (see for review Malina, 2004). They better reach and grasp objects compare to 6-month-old infants (Fagard et al., 2009; C. Newman et al., 2001). Motor skills develops a lot during the end of the first year and the second year of life, and by 19 months, infants walk by themselves and some even start to walk up stairs (see for review Malina, 2004). Because of this increasingly motor capacities, infants gradually explore more and more their environment and interact consequently with even more objects and more special partners, as they better move by themselves, relying less and less on their caregivers. It is noteworthy to highlight that being able to categorize objects in the environment might favor infants' interactions with different objects, helping them to discover statistical regularities between certain visual features and conceptual categories (e.g., animacy, function).

Instead of motor skills eliciting categorical saliency, the reverse might be further explored: saliency of categorical features allows infants to better explore objects in their environment by themselves. They will be surprised if an object they spontaneously recognized as inanimate start to move by itself; but they won't be so surprised to see the cat suddenly standing and heading to the

kitchen, where its dry food is, and might even be interested into following it. Although infants are already able to represent the animate and inanimate categories of objects, they appeared to start paying attention to the categories of objects when they start to increasingly move by themselves and interact with more and more objects.

In this thesis, I have focused on static images categorization, studying static features of images in the development of categorization; but in the world, one key feature of animate objects, is that they move by themselves, driven by an internal force, when inanimate objects require an external force to *be moved*. Several studies suggest that, early on, infants first and foremost rely on motion when exploring objects in their environment; indeed, already at birth, infants discriminate and prefer to look at biological motion compare to other type of motion (Bardi et al., 2011; Simion et al., 2008). This is also supported by previous results showing that 3- to 5-month-old infants infer goals to an inanimate object, a box, that has the key feature of animate, as it is self-propelled (Luo, 2011; Luo & Baillargeon, 2005).

7.2.2. The special case of the animate-inanimate categorization

All Chapters revealed the animate-inanimate categorization throughout development, from 4 months of age. In addition to Chapters 3-4, revealing animacy categorization in 4-, 6-, 8-, 10- and 19-month-olds' behavior, Chapter 5 and 6 additionally revealed that 4-month-old infants' brain automatically extract visual features diagnostic of the animate and inanimate categories, relying on another paradigm, and an even bigger set of stimuli. In these Chapters, I recorded infants' brain activity, using electroencephalography (EEG). Specifically, I used the frequency-tagging paradigm, to further study the animate-inanimate categorization in 4-month-old infants. The frequency-tagging paradigm aims at capturing an automatic and fast response of the brain (de Heering & Rossion, 2015; Peykarjou, 2022; Rossion et al., 2015), tagging a specific brain response, occurring each time the subject is presented with a specific stimulus, presented regularly. As the stimulation occurrence is regular, it is possible to capture the brain response in the EEG frequency spectrum. Indeed, a regular brain activity would be reflected by a peak of activity at the frequency of presentation of the tagged category, in the frequency domain of the EEG signal. However, this peak would not be visible if the brain response was always different between the images, as then, there would be no regular brain activity. Thus, such peak is diagnostic of a categorical response of the brain. In this case, the regular stimulation was for a category, meaning that I presented different exemplars of one category (e.g., the animate category) at

a specific frequency, within different exemplars of the other category (e.g., the inanimate category). I used a stimuli set that was designed to represent, as much as possible, the colossal variability present in the world, both within and between the animate and inanimate categories. In particular, the inanimate set of stimuli was, to my knowledge, the largest ever used. Until now, past research focused on the artificial, man-made categories (B. Elsner et al., 2013; Jeschonek et al., 2010; Peykarjou et al., 2023, 2024), or on the plant category (C. Elsner & Wertz, 2019; Wertz & Wynn, 2014a, 2014b), but not to both the natural and artificial categories at the same time, as part of the inanimate category. I considered big and small natural and artificial objects in the inanimate category, ensuring that any categorical response would correspond to the categorization of objects as either animate or inanimate objects.

Chapters 5 and 6 revealed a categorical response of objects as animate and inanimate, for 4-month-old infants. This response was captured over posterior electrodes, adding evidence for an automatic categorization of animate and inanimate objects based on visual features at 4 months of age, despite the huge variability, both within and between categories, and even when images were impoverished so that they were not recognizable anymore, but preserved lower-level visual features such as the color, contrast, luminance and spectral properties, although those images elicited smaller categorical response as compare to the intact images (Supplementary Results of Chapter 6).

Those results suggest that the first categorization infants represent (and rely on) is the animate-inanimate categorization, adding evidence to already existent evidences of global-level (i.e., superordinate-level) categorization being the first represented across development (Pauen, 2002; Peykarjou et al., 2024; Quinn & Johnson, 2000).

Being able to categorize animate and inanimate in our environment provides us, but also other species, with a good survival mechanism. Indeed, as categorization is a fast process, occurring as soon as human adults detect an object in a visual scene (Grill-Spector & Kanwisher, 2005), categorizing objects as animate and inanimate facilitates the attention to animate stimuli that can represent predators or preys. Such survival advantage might explain that this categorization is observed from early on in life (Spriet et al., 2022), in different species (see for review Freedman & Miller, 2008), especially in primates (Bell et al., 2009; Kiani et al., 2007; Kriegeskorte, Mur, Ruff, et al., 2008; see for a review Tompa & Sáry, 2010), influencing the organization of the ventral stream, this brain territory dedicated to decoding of visual objects (e.g., Carlson et al., 2013; Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008), but also influencing human behavior, as suggested by results of Chapters 3-4 in infancy, as well as results in Chapter 6 in adults. Studies further revealed that animacy appears as an

organizing principles for both grammar and discourse (Dahl, 2008), influence human memory (Bonin et al., 2014; Hagen & Laeng, 2017; Nairne et al., 2013, 2017), and visuo-attentional capacities (Guerrero & Calvillo, 2016; Lindh et al., 2019), or children interactions (LoBue et al., 2013).

7.3. Factors influencing the animate-inanimate categorization

To categorize objects into meaningful units (i.e., categories) facilitates every aspect of our life. Instead of having to store every little details for each object we have to interact with in our life, categories permit to store general, shared knowledge for groups of objects sharing similar properties, reducing the effort we have to do by a considerable amount. Categories can be formed based on an infinite number of properties – the most useful ones in our everyday life might be the functional properties of objects (e.g., toy, phone, food, pets...). However, when very little, or even no knowledge at all is available for an object, as it is the case the first time we encounter something new (e.g., virtual reality machine or simulation machine if we don't have one at home or at work), we are still able to explore it and interact with it accordingly, to a certain extent. Indeed, even though we don't know the object, or even recognize its basic category and function, we recognize features that indicate its superordinate category (i.e., whether it is an animal or an inanimate object, a natural or an artificial object). This is true for adults, but it is especially true for a newborn, that is constantly confronted to novel objects. One of the first exploration of objects we do, both in our everyday life as well as in our lifetime, is visual: we move our eyes and see the objects in our environment, and the features composing them, before touching them. As such, it is not so surprising that our visual system itself is already organized according to important broad categories (Caramazza & Shelton, 1998; Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Op de Beeck et al., 2008, 2019). Those categories nicely correspond to conceptual categories, such as the broad distinction between animate and inanimate objects, so that we can name those categories easily; yet, those are visual categories, categories that are based on the similarity of the visual features of the objects. We have seen that early on, human infants are already able to categorize objects, at least between animate and inanimate objects, as soon as 4 months of age. This categorization is based on the visual properties of the objects, as I mainly showed novel unknown objects to infants. Results of Chapter 5 show that this categorization between animate and inanimate objects is indeed especially recruiting posterior regions of infants' brain. Chapters 5 and 6 especially studied the factors influencing the animate-inanimate categorization. Such factors were

also investigated with the correlation between infants' behavior and the organization of objects as recorded in the adults' ventral stream (Chapters 3-4).

7.3.1. The animacy categorization is recruiting many features

Different features are processed in the adults' visual system, throughout the ventral stream, forming a hierarchy of the level of processing from posterior to anterior regions (Groen et al., 2017). The most posterior regions of the ventral stream process features that are referred to as low-level; those features correspond to the local measurements of the visual stimulus, such as the local orientations, spatial frequency, color or contrast of the object (Groen et al., 2017; Xie et al., 2022). In mid-level regions are processed the mid-level features, that correspond to more global measurement of bigger parts of the object, such as the global form (how curvy or rectilinear the object is) or the texture (Groen et al., 2017; Long et al., 2017). Finally, in the most anterior regions of the ventral stream are processed higher-level visual features, corresponding to the overall organization of the smaller parts of the object into the configuration of the entity, such as faces, plants and other objects (Groen et al., 2017). This is the place of the object recognition, allowing the identification of the object, and the semantic access to all information we have from it.

Previous research have demonstrated that earlier areas, so earlier level of processing of the object, such as mid-level features, already elicit in adults a categorization of the objects by animacy, even influencing humans' behavior (Long et al., 2017, 2018). Other results provided evidences that some features, such as the shape of the objects, might be more important than other features in the adults' representation of the categories, although not sufficient to fool adults' explicit recognition of objects' category (Bracci et al., 2019; Bracci & Op de Beeck, 2016; Chen et al., 2023). Chapter 6 aimed at studying the role of various levels of features in the animacy categorization in adults' brain as well as adults' behavior, and revealed that even low-level visual features, such as the power spectrum of the object, or the color, contrast and luminance, already elicited a categorization of the objects by animacy, both in adults and infants (Chapter 6 – Supplementary Results). Thus, information relevant for the animate-inanimate categorization are already extracted from lower-level visual features. However, unrecognizable stimuli that only contained lower- to mid-level visual features elicited smaller categorization than the recognizable stimuli preserving higher-level visual information. Furthermore, adults' behavior revealed that unrecognized images depicting only low-level features of the object (phase-scramble images) were not correctly categorized in a forced-choice task by adults, while images

containing more mid-level features were. Overall, these results suggest that the more visual features are available, the stronger the brain categorization by animacy, and the influence on humans' behavior.

Chapters 3 and 4 further suggested that young infants represent animate and inanimate categories by mainly recruiting higher-level visual features, as suggested by the correlation between infants' behavior and the most anterior part of the adults' ventral stream. Indeed, the anterior part of the adults' ventral stream is where the higher-level visual features are represented and extracted from the visual inputs. The fact that infants, when (and only when) representing the animate and inanimate categories, show a similar representation of the objects as compare to this anterior part suggests that infants' representation of the objects are mainly based on higher-level visual features, extracted in those anterior regions. As they grow older, infants' behavior did correlate with increasing portion of the ventral stream, including more posterior areas, suggesting that they now also recruit lower-level visual features in their representation of objects. While doing so, they also revealed a better distinctiveness of the animate-inanimate categories (Supplementary Results of Chapter 4), meaning that the difference between within-category dissimilarity and between-category dissimilarity increase with age, as infants better represent the within-category similarities and/or the differences between animate and inanimate. Overall, it suggests that representing the visual features of animacy (i.e., co-varying and diagnostic of the categories) already from the lower-level features largely improves the representation of animacy, but also that lower-level visual features are not yet eliciting categorical distinction in early infancy. This might reveal a potential tuning of lower-level visual features towards visual features diagnostic of a category. Such tuning might develop by statistical learning, after being exposed to many visual objects revealing the co-variance of lower-level visual features with categories, or by better connections between the different brain areas and increasing feedback processing from anterior regions to posterior regions in the ventral stream.

All those results converge together, and support the idea that the categorization of animate and inanimate objects is an additive process, in which the accumulation of visual information increases categorization performance, starting from the representation of higher-level visual features early in life and recruiting all level of features later on, optimizing the perception in classifying animate and inanimate stimuli from the early stages of visual processing.

7.3.2. Acceleration of visual object categorization throughout life

Chapter 5 showed that the animacy categorization is both fast, and automatic, accelerating throughout development. This categorization was especially recorded in posterior areas on the infants' brain, encompassing the occipital electrodes, adding evidence that it is indeed a visual categorization of the stimuli that is captured, as soon as 4 months of age. The speed of categorization did increase with development. Indeed, 4-month-old infants revealed categorization of objects when presented for at least 250ms (but not when presented for 166 ms), while 9-month-old infants revealed categorization of objects when images were presented for as few as 83ms. There is, in 5 months of development, an acceleration of the categorization processes by at least a factor of 3; 9-month-old infants are at least 3 times faster compared to 4-month-old infants. This process might still accelerate through life, and Chapter 5 revealed that adults were 7 times faster than 4-month-old infants, and possibly 2 to 3 times faster than 9-month-old infants in categorizing images as animate and inanimate. Indeed, results showed categorization of objects in adults with images flashed for 33ms. This acceleration of categorization might be linked to the previous observation that infants, at 4 to 6 months of age, revealed an animacy categorization based on mainly higher-level visual features, while older infants also recruited features extracted all along the adults' ventral stream, including lower-level features. The recruitment of more and more features in the animacy representation might be facilitated by the acceleration of information processing, in turns, accelerating the speed of categorization. Indeed, studies revealed that different visual features activate the ventral stream in different areas, from posterior to anterior regions; as doing so, the activation for the different features decays at a different speed, from faster decaying in posterior regions to slower decaying in anterior regions (Gao et al., 2020; Kiebel et al., 2008). Thus, being slower to integrate features would have an effect on the amount and the level of features still represented: only the features that are encoded in the most anterior regions of the ventral stream would still be available for the integration. Such theory is congruent with the observation that 4- and 6-month-old infants' representation of the animate and inanimate categories is mainly based on higher-level visual features. Then, with age, the information processing gets faster and faster, adding more and more visual features in the integration, allowing a faster categorization and a better representation of the world.

Noteworthy, this acceleration in the processing of categories echoes the one observed for attentional processes, from 4 to 9 months of age (Hochmann & Kouider, 2022). It is not clear whether attentional processes play a role in this automatic categorization of objects. However, previous studies

have found, in adults, that the perceptual awareness of the category is linked to the robustness of the categorical response of the brain (Rekow, Baudouin, Brochard, et al., 2022; Retter et al., 2020). It might be that only when the attentional system is fast enough, the categorization of objects accelerates as well.

Overall, being faster in categorizing objects in our environment facilitates our comprehension of the world; as growing up, we become more and more expert and fast in categorizing, thus recognizing, objects in our environment, improving our interaction with them, facilitating our choices to sit on the couch and not on the floor or to eat the delicious strawberries instead of spinach.

7.4. Open questions and future directions

Despite highlighting visual categorization influence in infants' behavior, especially the animate-inanimate categorization, and the role of higher-level visual features in the representation of objects, this thesis leaves open a number of questions. In particular, we began our investigations with 4-month-olds, thus begging the question of what happens earlier in development, particularly in neonates. Our study with preterm infants showed a major role of visual experience to increase the saliency of the animate-inanimate categories, but there are several reasons to believe that the representation of those categories pre-exist, are genetically predefined and may therefore be observed at birth. Several populations with radically different visual experience indeed exhibit the same broad organization of the ventral stream by animacy (macaques, congenitally blind humans), suggesting that this organization is independent from visual experience.

Another unanswered question is whether infants make conceptual inferences from the identification of the visual categories that we studied. Previous studies suggest that infants expect different things from animate or inanimate objects, such as being able to move by itself, have goals (Luo, 2011; Luo & Baillargeon, 2005; A. Woodward, 1998; A. L. Woodward, 1999), being non-hollowed (Kominsky et al., 2021; Setoh et al., 2013) and have social affiliation (Powell & Spelke, 2013, 2018) for animate, while inanimate are expected to follow physics principles, being continuous (Spelke, 1994), moving only by contact (Spelke, Phillips, et al., 1995) without intentionality (A. Woodward, 1998). But what knowledge infants extract from static images in those designs remain unclear. One can also ask how language interfere with the representation of visual categories; is it refining them, or is it only helping more conceptual categories to emerge? Although it appears that language invites infants to categorize objects that are designated by the same label (Balaban & Waxman, 1996; Ferry et al.,

2010, 2013; LaTourrette & Waxman, 2020, 2022; Waxman & Markow, 1995), it is not clear whether infants visual representation change when learning the name of objects they already represent, or whether a new category is formed, associating common visual features to a common name.

A similar question comes with the role of attentional processes in the representation of visual categories. Does the representation of categories require attention, automatically recruiting and using attentional resources? Is it only a coincidence that infants start to use visual categories in their visual exploration when they start to move by themselves? However, preterm infants' behavior, whose motor development can be delayed but visual categorization is not, suggest that the development of visual categories may be independent from motor development, although more studies are needed to investigate such link.

Another interesting question comes with the level of features in the animate-inanimate categorization. The animacy categorization is recorded in the adults' and infants' brain even from images only retaining lower-level visual features, while not influencing their behavior, and while other low-level features overshadowed the influence of categorical features in infants' behavior. Is there any advantage for our brain to be able to capture such an early categorization? Those early process might facilitate to flee a predator, preparing to run before we even be conscious about the presence of a predator in the environment. Such effect could be investigated by recording of cardiac rhythm or sudation, or of muscular activity.

Finally, another interesting question not addressed in this thesis is what happens if there is an early dysfunction of visual categorization processes? I expect that such impairments would have big consequences on cognition. An easy category impairment and its consequences to think of is the human face categorization. If infants are delayed in their representation of the human face category, they might not pay attention to social features when they were supposed to, and this might impact the development of social cognition. A very known disorder of social cognitive skills is the autism spectrum disorder (ASD); some evidence suggests that overall categories, not just the human face category, might be impacted in ASD subjects (Gastgeb & Strauss, 2012; C. R. Johnson & Rakison, 2006; Vanmarcke et al., 2016).

7.5. Conclusion

In conclusion, this work highlight the development of visual categories, starting with the broad animate-inanimate categorization, and its role in infants' behavior. I propose that being able, early in life, to already distinguish between what is animate and what is not helps infants to build expectations about novel objects they may encounter. It helps infants making sense of this numerous and varying stimulations they receive. Growing up, infants get better and faster at categorizing and recognizing objects, leading them to be able to interact with every objects accordingly, and to navigate properly, to take informed decisions, to think about objects, to live their life.

8. Bibliography

- Adams, R. J. (1995). Further Exploration of Human Neonatal Chromatic-Achromatic Discrimination. *Journal of Experimental Child Psychology*, 60(3), 344-360. <https://doi.org/10.1006/jecp.1995.1045>
- Adibpour, P., Dubois, J., & Dehaene-Lambertz, G. (2018). Right but not left hemispheric discrimination of faces in infancy. *Nature Human Behaviour*, 2(1), 67-79. <https://doi.org/10.1038/s41562-017-0249-4>
- Adibpour, P., & Hochmann, J.-R. (2023). Infants' understanding of the causal power of agents and tools. *Proceedings of the National Academy of Sciences*, 120(50), e2309669120. <https://doi.org/10.1073/pnas.2309669120>
- Adolph, K. E., & Hoch, J. E. (2019). Motor Development : Embodied, Embedded, Enculturated, and Enabling. *Annual Review of Psychology*, 70, 141-164. <https://doi.org/10.1146/annurev-psych-010418-102836>
- Anderson, J. R. (1991). The adaptive nature of Human categorization. *Psychological Review*, 98(3), 409-429.
- Arcaro, M. J., & Livingstone, M. S. (2017). A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife*, 6, e26196. <https://doi.org/10.7554/eLife.26196>
- Arcaro, M. J., Schade, P. F., & Livingstone, M. S. (2019). Body map proto-organization in newborn macaques. *Proceedings of the National Academy of Sciences*, 116(49), 24861-24871. <https://doi.org/10.1073/pnas.1912636116>
- Arterberry, M. E., & Bornstein, M. H. (2012). Categorization of Real and Replica Objects by 14- and 18-month-old Infants. *Infant behavior & development*, 35(3), 606-612. <https://doi.org/10.1016/j.infbeh.2012.04.007>
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48-53.
- Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences*, 9(3), 92-98. <https://doi.org/10.1016/j.tics.2005.01.003>
- Atkinson, J., Hood, B., Wattam-Bell, J., Anker, S., & Tricklebank, J. (1988). Development of Orientation Discrimination in Infancy. *Perception*, 17(5), 587-595. <https://doi.org/10.1068/p170587>
- Ayzenberg, V., & Lourenco, S. (2022). Perception of an object's global shape is best described by a model of skeletal structure in human infants. *eLife*, 11, e74943. <https://doi.org/10.7554/eLife.74943>
- Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, 9(1), 9359. <https://doi.org/10.1038/s41598-019-45268-y>
- Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing : Backward masking and natural scene categorisation. *Vision Research*, 45(11), 1459-1469. <https://doi.org/10.1016/j.visres.2005.01.004>
- Baillargeon, R., Li, J., Ng, W., & Yuan, S. (2008). An Account of Infants' Physical Reasoning. In A. Woodward & A. Needham (Éds.), *Learning and the Infant Mind* (p. 66-116). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195301151.003.0004>

- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110-118. <https://doi.org/10.1016/j.tics.2009.12.006>
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191-208. [https://doi.org/10.1016/0010-0277\(85\)90008-3](https://doi.org/10.1016/0010-0277(85)90008-3)
- Baillargeon, R., Wu, D., Yuan, S., Li, J., & Luo, Y. (2009). Young infants' expectations about self-propelled objects. In B. M. Hood & L. R. Santos (Éds.), *The Origins of Object Knowledge* (p. 285-352). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199216895.003.0012>
- Balaban, M. T., & Waxman, S. R. (1996). An examination of the factors underlying the facilitative effect of word phrases on object categorization in 9-month-old infants. *Proceedings of the 20th Boston University Conference on Language Development*, *1*, 483-493.
- Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth : Role of dynamic and configural properties. *Developmental Science*, *14*(2), 353-359. <https://doi.org/10.1111/j.1467-7687.2010.00985.x>
- Bayet, L., Zinszer, B. D., Reilly, E., Cataldo, J. K., Pruitt, Z., Cichy, R. M., Nelson, C. A., & Aslin, R. N. (2020). Temporal dynamics of visual representations in the infant brain. *Developmental Cognitive Neuroscience*, *45*, 100860. <https://doi.org/10.1016/j.dcn.2020.100860>
- Bayne, T., Brainard, D., Byrne, R. W., Chittka, L., Clayton, N., Heyes, C., Mather, J., Ölveczky, B., Shadlen, M., Suddendorf, T., & Webb, B. (2019). What is cognition? *Current Biology*, *29*(13), R608-R615. <https://doi.org/10.1016/j.cub.2019.05.044>
- Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*(2), 105-141. [https://doi.org/10.1016/0010-0277\(96\)00706-8](https://doi.org/10.1016/0010-0277(96)00706-8)
- Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B. H., & Ungerleider, L. G. (2009). Object Representations in the Temporal Cortex of Monkeys and Humans as Revealed by Functional Magnetic Resonance Imaging. *Journal of Neurophysiology*, *101*(2), 688-700. <https://doi.org/10.1152/jn.90657.2008>
- Bender, A., Voytek, B., & Schaworonkow, N. (2023). Resting-state is not enough : Alpha and mu rhythms change shape across development, but lack diagnostic sensitivity. *bioRxiv*. <https://doi.org/10.1101/2023.10.13.562301>
- Bhatt, R. S., Hock, A., White, H., Jubran, R., & Galati, A. (2016). The development of body structure knowledge in infancy. *Child Development Perspectives*, *10*(1), 45-52. <https://doi.org/10.1111/cdep.12162>
- Biederman, I. (1987). Recognition-by-Components : A Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115-147.
- Biran, I., & Coslett, H. B. (2003). Visual agnosia. *Current Neurology and Neuroscience Reports*, *3*(6), 508-512. <https://doi.org/10.1007/s11910-003-0055-4>
- Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., & Love, B. C. (2020). Measures of neural similarity. *Computational Brain & Behavior*, *3*(4), 369-383. <https://doi.org/10.1007/s42113-019-00068-5>
- Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294-328. [https://doi.org/10.1016/0022-0965\(83\)90085-1](https://doi.org/10.1016/0022-0965(83)90085-1)

- Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis : Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, *44*(4), 388-426. <https://doi.org/10.1006/cogp.2002.0779>
- Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates : Further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370-382. <https://doi.org/10.3758/s13421-013-0368-8>
- Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children : Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology*, *46*(2), 350-365. <https://doi.org/10.1037/a0018411>
- Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color Vision and Hue Categorization in Young Human Infants. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 115-129.
- Bornstein, M., Kessen, W., & Weiskopf, S. (1976). The categories of hue in infancy. *Science*, *191*(4223), 201-202. <https://doi.org/10.1126/science.1246610>
- Bosworth, R. G., & Dobkins, K. R. (2009). Chromatic and luminance contrast sensitivity in fullterm and preterm infants. *Journal of Vision*, *9*(13), 1-16. <https://doi.org/10.1167/9.13.15>
- Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *The Journal of Neuroscience*, *36*(2), 432-444. <https://doi.org/10.1523/JNEUROSCI.2314-15.2016>
- Bracci, S., & Op De Beeck, H. P. (2023). Understanding Human Object Vision : A Picture Is Worth a Thousand Representations. *Annual Review of Psychology*, *74*(1), 113-135. <https://doi.org/10.1146/annurev-psych-032720-041031>
- Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. P. (2019). The Ventral Visual Pathway Represents Animal Appearance over Animacy, Unlike Human Behavior and Deep Neural Networks. *The Journal of Neuroscience*, *39*(33), 6513-6525. <https://doi.org/10.1523/JNEUROSCI.1714-18.2019>
- Braddick, O. J., Atkinson, J., & Wattam-Bell, J. R. (1986). Development of the discrimination of spatial phase in infancy. *Vision Research*, *26*(8), 1223-1239. [https://doi.org/10.1016/0042-6989\(86\)90103-3](https://doi.org/10.1016/0042-6989(86)90103-3)
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436. <https://doi.org/10.1163/156856897X00357>
- Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science*, *9*(6), F59-F64. <https://doi.org/10.1111/j.1467-7687.2006.00530.x>
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (2017). *A Study of Thinking*. Routledge. <https://doi.org/10.4324/9781315083223>
- Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., Baldo, E., & Vallortigara, G. (2019). Cortical route for facelike pattern processing in human newborns. *Proceedings of the National Academy of Sciences*, *116*(10), 4625-4630. <https://doi.org/10.1073/pnas.1812419116>
- Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy*, *1*(2), 149-219.

- Candy, T. R. (2019). The Importance of the Interaction Between Ocular Motor Function and Vision During Human Infancy. *Annual Review of Vision Science*, 5(1), 201-221. <https://doi.org/10.1146/annurev-vision-091718-014741>
- Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, 20(3-6), 213-261. <https://doi.org/10.1080/02643290244000266>
- Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge : The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8), 354-361. [https://doi.org/10.1016/S1364-6613\(03\)00159-1](https://doi.org/10.1016/S1364-6613(03)00159-1)
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain : The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 10(1), 1-34.
- Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.
- Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*, 26(1), 132-142. https://doi.org/10.1162/jocn_a_00476
- Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The Emergence of Semantic Meaning in the Ventral Temporal Pathway. *Journal of Cognitive Neuroscience*, 26(1), 120-131. https://doi.org/10.1162/jocn_a_00458
- 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>
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, 63(4), i. <https://doi.org/10.2307/1166214>
- Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of Parent Talk on the Development of Preterm Infant Vocalizations. *Pediatrics*, 128(5), 910-916. <https://doi.org/10.1542/peds.2011-0609>
- Cavallini, A., Fazzi, E., Viviani, V., Astori, M. G., Zaverio, S., Bianchi, P. E., & Lanzi, G. (2002). Visual acuity in the first two years of life in healthy term newborns : An experience with the Teller acuity cards. *Functional Neurology*, 17(2), 87-92.
- Celik, H. I., Elbasan, B., Gucuyener, K., Kayihan, H., & Huri, M. (2018). Investigation of the Relationship Between Sensory Processing and Motor Development in Preterm Infants. *The American Journal of Occupational Therapy*, 72(1), 7201195020p1-7201195020p7. <https://doi.org/10.5014/ajot.2018.026260>
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2(10), 913-919. <https://doi.org/10.1038/13217>
- Chatterjee, S., & Callaway, E. M. (2003). Parallel colour-opponent pathways to primary visual cortex. *Nature*, 426(6967), 668-671. <https://doi.org/10.1038/nature02167>
- Chen, C.-Y., Leys, G., Bracci, S., & Op De Beeck, H. (2023). The representational dynamics of the animal appearance bias in human visual cortex are indicative of fast feedforward processing. *Imaging Neuroscience*, 1, 1-26. https://doi.org/10.1162/imag_a_00006
- 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>

- Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The Role of Visual and Semantic Properties in the Emergence of Category-Specific Patterns of Neural Response in the Human Brain. *Eneuro*, 3(4), ENEURO.0158-16.2016. <https://doi.org/10.1523/ENEURO.0158-16.2016>
- Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of Neurophysiology*, 117(1), 388-402. <https://doi.org/10.1152/jn.00569.2016>
- Cohen, M. X. (2017). Where Does EEG Come From and What Does It Mean? *Trends in Neurosciences*, 40(4), 208-218. <https://doi.org/10.1016/j.tins.2017.02.004>
- 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>
- Dahl, Ö. (2008). Animacy and egophoricity : Grammar, ontology and phylogeny. *Lingua*, 118(2), 141-150. <https://doi.org/10.1016/j.lingua.2007.02.008>
- Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI : Practices and pitfalls: representational analysis using fMRI. *Annals of the New York Academy of Sciences*, 1296(1), 108-134. <https://doi.org/10.1111/nyas.12156>
- Davis, T., & Poldrack, R. A. (2014). Quantifying the Internal Structure of Categories Using a Neural Typicality Measure. *Cerebral Cortex*, 24(7), 1720-1737. <https://doi.org/10.1093/cercor/bht014>
- Dawkins, M. B., Ting, F., Stavans, M., & Baillargeon, R. (2020). Early Moral Cognition : A Principle-Based Approach. In *The cognitive neurosciences VI* (Vol. 67, p. 7-16).
- De Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy : A review. *International Journal of Psychophysiology*, 51(1), 45-58. [https://doi.org/10.1016/S0167-8760\(03\)00152-1](https://doi.org/10.1016/S0167-8760(03)00152-1)
- Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, 8, 1-10. <https://doi.org/10.1038/ncomms13995>
- de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, 4, e06564. <https://doi.org/10.7554/eLife.06564>
- Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, 47(6), 1499-1503. <https://doi.org/10.1037/a0025659>
- Delorme, A., & Makeig, S. (2004). EEGLAB : An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9-21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- DeMasi, A., Schoneveld, E., & Berger, S. E. (2023). Three decades of infant motor development : Cohort effects in motor skill onsets. *Developmental Psychology*. <https://doi.org/10.1037/dev0001628>
- Deoni, S. C. L., Mercure, E., Blasi, A., Gasston, D., Thomson, A., Johnson, M., Williams, S. C. R., & Murphy, D. G. M. (2011). Mapping infant brain myelination with magnetic resonance imaging. *Journal of Neuroscience*, 31(2), 784-791. <https://doi.org/10.1523/JNEUROSCI.2106-10.2011>
- Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology*, 43(5), 1227-1238. <https://doi.org/10.1037/0012-1649.43.5.1227>

- Deza, A., Chen, Y.-C., Long, B., & Konkle, T. (2019). Accelerated Texforms : Alternative Methods for Generating Unrecognizable Object Images with Preserved Mid-Level Features. *2019 Conference on Cognitive Computational Neuroscience*. 2019 Conference on Cognitive Computational Neuroscience, Berlin, Germany. <https://doi.org/10.32470/CCN.2019.1412-0>
- Di Lorenzo, R., van den Boomen, C., Kemner, C., & Junge, C. (2020). Charting development of ERP components on face-categorization : Results from a large longitudinal sample of infants. *Developmental Cognitive Neuroscience*, 100840. <https://doi.org/10.1016/j.dcn.2020.100840>
- Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants : A review and comparison of behavioral and electrophysiological studies. *Vision Research*, 18(11), 1469-1483. [https://doi.org/10.1016/0042-6989\(78\)90001-9](https://doi.org/10.1016/0042-6989(78)90001-9)
- Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex*, 16(10), 1453-1461. <https://doi.org/10.1093/cercor/bhj086>
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, 293(5539), 2470-2473. <https://doi.org/10.1126/science.1063414>
- Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Hüppi, P. S., & Hertz-Pannier, L. (2014). The early development of brain white matter : A review of imaging studies in fetuses, newborns and infants. *Neuroscience*, 276, 48-71. <https://doi.org/10.1016/j.neuroscience.2013.12.044>
- Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, 14(1-2), 1-52. [https://doi.org/10.1016/0301-0511\(82\)90016-3](https://doi.org/10.1016/0301-0511(82)90016-3)
- Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, 65, 903-917.
- Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, 58, 418-431.
- Elhamiasl, M., Sanches Braga Figueira, J., Barry-Anwar, R., Pestana, Z., Keil, A., & Scott, L. S. (2023). The emergence of the EEG dominant rhythm across the first year of life. *Cerebral Cortex*, bhad425. <https://doi.org/10.1093/cercor/bhad425>
- Elizabeth S. Spelke, E. P. B. & Amy E. Skerry. (2013). Core social cognition. In M. R. Banaji & S. A. Gelman (Éds.), *Navigating the Social World* (p. 11-16). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199890712.003.0003>
- Elsner, B., Jeschonek, S., & Pauen, S. (2013). Event-related potentials for 7-month-olds' processing of animals and furniture items. *Developmental Cognitive Neuroscience*, 3, 53-60. <https://doi.org/10.1016/j.dcn.2012.09.002>
- Elsner, C., & Wertz, A. E. (2019). The seeds of social learning : Infants exhibit more social looking for plants than other object types. *Cognition*, 183, 244-255. <https://doi.org/10.1016/j.cognition.2018.09.016>
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601. <https://doi.org/10.1038/33402>
- Fagard, J., Spelke, E., & von Hofsten, C. (2009). Reaching and grasping a moving object in 6-, 8-, and 10-month-old infants : Laterality and performance. *Infant Behavior and Development*, 32(2), 137-146. <https://doi.org/10.1016/j.infbeh.2008.12.002>

- Fantz, R. L. (1964). Visual Experience in Infants : Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science*, *146*(3644), 668-670. <https://doi.org/10.1126/science.146.3644.668>
- Farah, M. J. (2004). *Visual agnosia* (2nd ed). MIT Press.
- Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli : Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245-17250. <https://doi.org/10.1073/pnas.0502205102>
- Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-Month-Old Infants : An Advantage of Words Over Tones. *Child Development*, *81*(2), 472-479. <https://doi.org/10.1111/j.1467-8624.2009.01408.x>
- Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences*, *110*(38), 15231-15235. <https://doi.org/10.1073/pnas.1221166110>
- Fielder, A. R., Moseley, M. J., & Ng, Y. K. (1988). The immature visual system and premature birth. *British Medical Bulletin*, *44*(4), 1093-1118. <https://doi.org/10.1093/oxfordjournals.bmb.a072291>
- Fox, R., Aslin, R. N., Shea, S. L., & Dumais, S. T. (1980). Stereopsis in Human Infants. *Science*, *207*(4428), 323-324. <https://doi.org/10.1126/science.7350666>
- Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349-377. <https://doi.org/10.1348/0261510041552738>
- Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization : Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews*, *19*.
- French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The Role of Bottom-Up Processing in Perceptual Categorization by 3- to 4-Month-Old Infants : Simulations and Data. *Journal of Experimental Psychology: General*, *133*(3), 382-397. <https://doi.org/10.1037/0096-3445.133.3.382>
- Freschl, J., Melcher, D., Carter, A., Kaldy, Z., & Blaser, E. (2021). Seeing a Page in a Flipbook : Shorter Visual Temporal Integration Windows in 2-Year-Old Toddlers with Autism Spectrum Disorder. *Autism Research*, *14*(5), 946-958. <https://doi.org/10.1002/aur.2430>
- Fuchigami, T., Okubo, O., Ejiri, K., Fujita, Y., Kohira, R., Noguchi, Y., Fuchigami, S., Hiyoshi, K., Nishimura, A., & Haradag, K. (1995). Developmental changes in P300 wave elicited during two different experimental conditions. *Pediatric Neurology*, *13*(1), 25-28. [https://doi.org/10.1016/0887-8994\(95\)00086-U](https://doi.org/10.1016/0887-8994(95)00086-U)
- Fuentefria, R. do N., Silveira, R. C., & Procianoy, R. S. (2017). Motor development of preterm infants assessed by the Alberta Infant Motor Scale : Systematic review article. *Jornal de Pediatria*, *93*(4), 328-342.
- Gao, R., Van Den Brink, R. L., Pfeffer, T., & Voytek, B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *eLife*, *9*, e61277. <https://doi.org/10.7554/eLife.61277>
- Gastgeb, H. Z., & Strauss, M. S. (2012). Categorization in ASD : The Role of Typicality and Development. *Perspectives on Language Learning and Education*, *19*(2). <https://doi.org/10.1044/ll19.2.66>
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183-209. [https://doi.org/10.1016/0010-0277\(86\)90034-X](https://doi.org/10.1016/0010-0277(86)90034-X)

- Gelman, S. A., & Meyer, M. (2011). Child categorization. *WIREs Cognitive Science*, 2(1), 95-105. <https://doi.org/10.1002/wcs.96>
- Gentile, F., & Rossion, B. (2014). Temporal frequency tuning of cortical face-sensitive areas for individual face perception. *NeuroImage*, 90, 256-265. <https://doi.org/10.1016/j.neuroimage.2013.11.053>
- Gerhardstein, P., Shroff, G., Dickerson, K., & Adler, S. A. (2009). The development of object recognition through infancy. In B. C. Glenyn & R. P. Zini (Éds.), *New directions in developmental psychobiology*. Nova Science Publishers.
- Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, 39(1), 1-42.
- Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in Human infants and adults. *Journal of Cognitive Neuroscience*, 17(8), 1328-1340. <https://doi.org/10.1162/0898929055002481>
- Gliga, T., & Dehaene-Lambertz, G. (2007). Development of a view-invariant representation of the human head. *Cognition*, 102(2), 261-288. <https://doi.org/10.1016/j.cognition.2006.01.004>
- Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science*, 15, 885-894.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20-26.
- Grill-Spector, K. (2003). *The functional organization of the ventral visual pathway and its relationship to object recognition*. 169-194.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual Recognition : As Soon as You Know It Is There, You Know What It Is. *Psychological Science*, 16(2), 152-160. <https://doi.org/10.1111/j.0956-7976.2005.00796.x>
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, 6(4), 316-328. [https://doi.org/10.1002/\(SICI\)1097-0193\(1998\)6:4<316::AID-HBM9>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0193(1998)6:4<316::AID-HBM9>3.0.CO;2-6)
- Grill-Spector, K., & Malach, R. (2004). The Human visual cortex. *Annual Review of Neuroscience*, 27(1), 649-677. <https://doi.org/10.1146/annurev.neuro.27.070203.144220>
- 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 B: Biological Sciences*, 372(1714), 20160102. <https://doi.org/10.1098/rstb.2016.0102>
- Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, 188, 668-679. <https://doi.org/10.1016/j.neuroimage.2018.12.046>
- Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2019). Untangling featural and conceptual object representations. *NeuroImage*, 202, 116083. <https://doi.org/10.1016/j.neuroimage.2019.116083>

- 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
- Grossmann, T., Gliga, T., Johnson, M. H., & Mareschal, D. (2009). *The Neural Basis of Perceptual Category Learning in Human Infants*. 21(12), 12.
- Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-Matter Connectivity between Face-Responsive Regions in the Human Brain. *Cerebral Cortex*, 22(7), 1564-1576. <https://doi.org/10.1093/cercor/bhr226>
- Guerrero, G., & Calvillo, D. P. (2016). Animacy increases second target reporting in a rapid serial visual presentation task. *Psychonomic Bulletin & Review*, 23(6), 1832-1838. <https://doi.org/10.3758/s13423-016-1040-7>
- Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of Spatial Contrast Sensitivity from Infancy to Adulthood : Psychophysical Data. *Optometry and Vision Science*, 74(10), 785-789.
- Gwiazda, J., Brill, S., Mohindra, I., & Held, R. (1978). Infant visual acuity and its meridional variation. *Vision Research*, 18(11), 1557-1564. [https://doi.org/10.1016/0042-6989\(78\)90011-1](https://doi.org/10.1016/0042-6989(78)90011-1)
- Gwiazda, J., Brill, S., Mohindra, I., & Held, R. (1980). Preferential Looking Acuity in Infants from Two to Fifty-Eight Weeks of Age. *Optometry and Vision Science*, 57(7), 428.
- Hagen, T., & Laeng, B. (2017). Animals Do Not Induce or Reduce Attentional Blinking, But They Are Reported More Accurately in a Rapid Serial Visual Presentation Task. *I-Perception*, 8(5), 204166951773554. <https://doi.org/10.1177/2041669517735542>
- Hainline, L. (1998). Summary and Commentary : Eye Movements, Attention and Development. In *Cognitive Neuroscience of Attention*. Psychology Press.
- Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing : Face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, 19(3), 1180-1193. [https://doi.org/10.1016/S1053-8119\(03\)00076-4](https://doi.org/10.1016/S1053-8119(03)00076-4)
- Hamlin, J. K. (2013). Moral Judgment and Action in Preverbal Infants and Toddlers : Evidence for an Innate Moral Core. *Current Directions in Psychological Science*, 22(3), 186-193. <https://doi.org/10.1177/0963721412470687>
- Happé, F., & Frith, U. (2006). The Weak Coherence Account : Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5-25. <https://doi.org/10.1007/s10803-005-0039-0>
- Harnad, S. (2005). To cognize is to categorize : Cognition is categorization. In *Handbook of Categorization in Cognitive Science* (p. 19-43). Elsevier. <https://doi.org/10.1016/B978-008044612-7/50056-1>
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404-416. <https://doi.org/10.1016/j.neuron.2011.08.026>
- Haxby, J. V., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B., & Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Nat. Acad. Sci. USA*, 88, 1621-1625.

- Hochmann, J.-R. (2013). Word frequency, function words and the second gavagai problem. *Cognition*, 128(1), 13-25. <https://doi.org/10.1016/j.cognition.2013.02.014>
- Hochmann, J.-R., & Kouider, S. (2022). Acceleration of information processing en route to perceptual awareness in infancy. *Current Biology*, 32(5), 1206-1210.e3. <https://doi.org/10.1016/j.cub.2022.01.029>
- Howland, H. C., Atkinson, J., Braddick, O., & French, J. (1978). Infant Astigmatism Measured by Photorefractometry. *Science*, 202(4365), 331-333. <https://doi.org/10.1126/science.694540>
- Huang, L. (2020). Space of preattentive shape features. *Journal of Vision*, 20(4), 10. <https://doi.org/10.1167/jov.20.4.10>
- Hubel, D. H., & Wiesel, T. N. (1998). Early Exploration of the Visual Cortex. *Neuron*, 20(3), 401-412. [https://doi.org/10.1016/S0896-6273\(00\)80984-8](https://doi.org/10.1016/S0896-6273(00)80984-8)
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69-95.
- Jagadeesh, A. V., & Livingstone, M. (2024). Texture bias in primate ventral visual cortex. *In ICLR 2024 Workshop on Representational Alignment*.
- James, W. (1890). *The Principles of Psychology: Vol. II*.
- Jandó, G., Mikó-Baráth, E., Markó, K., Hollódy, K., Török, B., & Kovacs, I. (2012). Early-onset binocularity in preterm infants reveals experience-dependent visual development in humans. *Proceedings of the National Academy of Sciences*, 109(27), 11049-11052. <https://doi.org/10.1073/pnas.1203096109>
- Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants? *Brain and Development*, 32(10), 863-871. <https://doi.org/10.1016/j.braindev.2009.11.010>
- Johnson, C. R., & Rakison, D. H. (2006). Early Categorization of Animate/Inanimate Concepts in Young Children with Autism. *Journal of Developmental and Physical Disabilities*, 18(2), 73-89. <https://doi.org/10.1007/s10882-006-9007-7>
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2), 1-19. [https://doi.org/10.1016/0010-0277\(91\)90045-6](https://doi.org/10.1016/0010-0277(91)90045-6)
- 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>
- Kabdebon, C., Fló, A., de Heering, A., & Aslin, R. (2022). The power of rhythms : How steady-state evoked responses reveal early neurocognitive development. *NeuroImage*, 254, 119150. <https://doi.org/10.1016/j.neuroimage.2022.119150>
- Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, 115(4), 2246-2250. <https://doi.org/10.1152/jn.01074.2015>
- Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *Journal of Neuroscience*, 36(41), 10522-10528. <https://doi.org/10.1523/JNEUROSCI.1385-16.2016>

- Kamps, F. S., Hendrix, C. L., Brennan, P. A., & Dilks, D. D. (2020). Connectivity at the origins of domain specificity in the cortical face and place networks. *Proceedings of the National Academy of Sciences*, 1-7. <https://doi.org/10.1073/pnas.1911359117>
- 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*, 17(11), 4302-4311.
- Kellman, P. J., & Arterberry, M. E. (2007). Infant Visual Perception. In W. Damon & R. M. Lerner (Éds.), *Handbook of Child Psychology* (p. chpsy0203). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0203>
- 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>
- Kheradpisheh, S. R., Ghodrati, M., Ganjtabesh, M., & Masquelier, T. (2016). Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition. *Scientific Reports*, 6(1), 32672. <https://doi.org/10.1038/srep32672>
- 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. <https://doi.org/10.1152/jn.00024.2007>
- Kiat, J. E., Luck, S. J., Beckner, A. G., Hayes, T. R., Pomaranski, K. I., Henderson, J. M., & Oakes, L. M. (2021). Linking patterns of infant eye movements to a neural network model of the ventral stream using representational similarity analysis. *Developmental Science*. <https://doi.org/10.1111/desc.13155>
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A Hierarchy of Time-Scales and the Brain. *PLoS Computational Biology*, 4(11), e1000209. <https://doi.org/10.1371/journal.pcbi.1000209>
- Knoblauch, K., Vital-Durand, F., & Barbur, J. L. (2001). Variation of chromatic sensitivity across the life span. *Vision Research*, 41(1), 23-36. [https://doi.org/10.1016/S0042-6989\(00\)00205-4](https://doi.org/10.1016/S0042-6989(00)00205-4)
- Kominsky, J. F., Shafto, P., & Bonawitz, E. (2021). “There’s something inside” : Children’s intuitions about animate agents. *PLOS ONE*, 16(5), e0251081. <https://doi.org/10.1371/journal.pone.0251081>
- Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, 33(25), 10235-10242. <https://doi.org/10.1523/JNEUROSCI.0983-13.2013>
- Konkle, T., & Oliva, A. (2012a). A familiar-size stroop effect : Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 561-569. <https://doi.org/10.1037/a0028294>
- Konkle, T., & Oliva, A. (2012b). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, 74(6), 1114-1124. <https://doi.org/10.1016/j.neuron.2012.04.036>
- Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, 32, 1-10.
- Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., De Gardelle, V., Christophe, A., Dehaene, S., & Dehaene-Lambertz, G. (2013). A Neural Marker of Perceptual Consciousness in Infants. *Science*, 340(6130), 376-380. <https://doi.org/10.1126/science.1232509>

- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The Social Sense : Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330(6012), 1830-1834. <https://doi.org/10.1126/science.1190792>
- Kramer, L. E., Konkle, T., Chen, Y.-C., Long, B., & Cohen, M. R. (2023). Contributions of early and mid-level visual cortex to high-level object categorization. *bioRxiv*, 2023.05.31.541514. <https://doi.org/10.1101/2023.05.31.541514>
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 1-28. <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>
- Kubilius, J., Bracci, S., & Op De Beeck, H. P. (2016). Deep Neural Networks as a Computational Model for Human Shape Sensitivity. *PLOS Computational Biology*, 12(4), e1004896. <https://doi.org/10.1371/journal.pcbi.1004896>
- Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition*, 94(1), 95-103. <https://doi.org/10.1016/j.cognition.2004.02.007>
- Kwon, M.-K., Luck, S. J., & Oakes, L. M. (2014). Visual Short-Term Memory for Complex Objects in 6- and 8-month-old Infants. *Child development*, 85(2), 564-577. <https://doi.org/10.1111/cdev.12161>
- LaTourrette, A., & Waxman, S. R. (2020). Naming guides how 12-month-old infants encode and remember objects. *Proceedings of the National Academy of Sciences*, 117(35), 21230-21234. <https://doi.org/10.1073/pnas.2006608117>
- LaTourrette, A., & Waxman, S. R. (2022). Sparse labels, no problems : Infant categorization under challenging conditions. *Child Development*, 93(6), 1903-1911. <https://doi.org/10.1111/cdev.13818>
- Lawson, R. P., Friston, K. J., & Rees, G. (2015). A more precise look at context in autism. *Proceedings of the National Academy of Sciences*, 112(38), E5226. <https://doi.org/10.1073/pnas.1514212112>
- Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., & Baudouin, J. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Developmental Science*, 23(2). <https://doi.org/10.1111/desc.12877>
- Lewis, T. L., Maurer, D., & Kay, D. (1978). Newborns' central vision : Whole or hole? *Journal of Experimental Child Psychology*, 26(1), 193-203. [https://doi.org/10.1016/0022-0965\(78\)90119-4](https://doi.org/10.1016/0022-0965(78)90119-4)
- Li, S. P. D., & Bonner, M. F. (2020). Curvature as an Organizing Principle of Mid-level Visual Representation : A Semantic-preference Mapping Approach. *In NeurIPS 2020 Workshop SVRHM*, 13.
- Libertus, K., Gibson, J., Hidayatallah, N. Z., Hirtle, J., Adcock, R. A., & Needham, A. (2013). Size matters : How age and reaching experiences shape infants' preferences for different sized objects. *Infant Behavior and Development*, 36(2), 189-198. <https://doi.org/10.1016/j.infbeh.2013.01.006>

- Libertus, K., & Hauf, P. (2017). Editorial : Motor Skills and Their Foundational Role for Perceptual, Social, and Cognitive Development. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00301>
- Lieber, J. D., Oleskiw, T. D., Simoncelli, E. P., & Movshon, J. A. (2024). Responses of neurons in macaque V4 to object and texture images. *bioRxiv*. <https://doi.org/10.1101/2024.02.20.581273>
- 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. *Nature Communications*, 10(1), 4106. <https://doi.org/10.1038/s41467-019-12135-3>
- Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia*, 52, 57-72. <https://doi.org/10.1016/j.neuropsychologia.2013.10.022>
- Livingstone, M. S., Vincent, J. L., Arcaro, M. J., Srihasam, K., Schade, P. F., & Savage, T. (2017). Development of the macaque face-patch system. *Nature Communications*, 8(1), 14897. <https://doi.org/10.1038/ncomms14897>
- Lloyd, B. B., & Rosch, E. (1978). *Cognition and categorization*. Routledge.
- LoBue, V., Bloom Pickard, M., Sherman, K., Axford, C., & DeLoache, J. S. (2013). Young children's interest in live animals : *Young children's interest in animals*. *British Journal of Developmental Psychology*, 31(1), 57-69. <https://doi.org/10.1111/j.2044-835X.2012.02078.x>
- Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition*, 168, 234-242. <https://doi.org/10.1016/j.cognition.2017.06.025>
- Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, 17(6), 20. <https://doi.org/10.1167/17.6.20>
- 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*, 115(38), E9015-E9024. <https://doi.org/10.1073/pnas.1719616115>
- Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent : Three-month-olds attribute goals to a non-human agent. *Developmental Science*, 14(2), 453-460. <https://doi.org/10.1111/j.1467-7687.2010.00995.x>
- Luo, Y., & Baillargeon, R. (2005). Can a Self-Propelled Box Have a Goal? : Psychological Reasoning in 5-Month-Old Infants. *Psychological Science*, 16(8), 601-608. <https://doi.org/10.1111/j.1467-9280.2005.01582.x>
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, 58(4), 441-486. <https://doi.org/10.1016/j.cogpsych.2008.11.001>
- Magri, C., Konkle, T., & Caramazza, A. (2020). The contribution of object size, manipulability, and stability on neural responses to inanimate objects. *bioRxiv*, 42. <https://doi.org/10.1101/2020.11.22.393397>
- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the Human brain does not require visual experience. *Neuron*, 63(3), 397-405. <https://doi.org/10.1016/j.neuron.2009.07.012>

- Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories : A cognitive neuropsychological perspective. *Annual Review of Psychology*, 60(1), 27-51. <https://doi.org/10.1146/annurev.psych.60.110707.163532>
- 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*, 92(18), 8135-8139. <https://doi.org/10.1073/pnas.92.18.8135>
- Malina, R. M. (2004). Motor Development during Infancy and Early Childhood : Overview and Suggested Directions for Research. *International Journal of Sport and Health Science*, 2, 50-66. <https://doi.org/10.5432/ijshs.2.50>
- Marcinowski, E. C., Tripathi, T., Hsu, L., Westcott McCoy, S., & Dusing, S. C. (2019). Sitting skill and the emergence of arms-free sitting affects the frequency of object looking and exploration. *Developmental Psychobiology*, 61(7), 1035-1047. <https://doi.org/10.1002/dev.21854>
- Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, 5(10), 443-450. [https://doi.org/10.1016/S1364-6613\(00\)01752-6](https://doi.org/10.1016/S1364-6613(00)01752-6)
- Marinović, V., Hoehl, S., & Pauen, S. (2014). Neural correlates of human–animal distinction : An ERP-study on early categorical differentiation with 4- and 7-month-old infants and adults. *Neuropsychologia*, 60, 60-76. <https://doi.org/10.1016/j.neuropsychologia.2014.05.013>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Markson, L., & Spelke, E. S. (2006). Infants' Rapid Learning About Self-Propelled Objects. *Infancy*, 9(1), 45-71. https://doi.org/10.1207/s15327078in0901_3
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 200(1140), 269-294.
- Martin, A. (2007). The Representation of Object Concepts in the Brain. *Annual Review of Psychology*, 58(1), 25-45. <https://doi.org/10.1146/annurev.psych.57.102904.190143>
- Mattioni, S., Rezk, M., Battal, C., Bottini, R., Cuculiza Mendoza, K. E., Oosterhof, N. N., & Collignon, O. (2020). Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *eLife*, 9, e50732. <https://doi.org/10.7554/eLife.50732>
- 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](https://doi.org/10.1016/S1364-6613(02)01903-4)
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the Human fusiform gyrus. *Journal of Cognitive Neuroscience*, 9(5), 605-610. <https://doi.org/10.1162/jocn.1997.9.5.605>
- McCarty, M. E., Clifton, R. K., Ashmead, D. H., Lee, P., & Goubet, N. (2001). How infants use vision for grasping objects. *Child Development*, 72(4), 973-987. <https://doi.org/10.1111/1467-8624.00329>
- Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology*, 12(4), 496-522. [https://doi.org/10.1016/0010-0285\(80\)90018-3](https://doi.org/10.1016/0010-0285(80)90018-3)
- Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32(1), 89-115. <https://doi.org/10.1146/annurev.ps.32.020181.000513>

- Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, 6(1), 57-77. [https://doi.org/10.1016/0166-4328\(82\)90081-X](https://doi.org/10.1016/0166-4328(82)90081-X)
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision : Two cortical pathways. *Trends in Neurosciences*, 6, 414-417.
- Mohindra, I., Held, R., Gwiazda, J., & Brill, S. (1978). Astigmatism in Infants. *Science*, 202, 329-331. <https://doi.org/10.1126/science.694539>
- Morison, V., & Slater, A. (1985). Contrast and Spatial Frequency Components in Visual Preferences of Newborns. *Perception*, 14(3), 345-348. <https://doi.org/10.1068/p140345>
- Mottron, L., & Burack, J. A. (2001). Enhanced Perceptual Functioning in the Development of Autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Éds.), *The development of autism : Perspectives from theory and research* (p. 131-148). Routledge.
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, 61(2), 63-86. <https://doi.org/10.1016/j.cogpsych.2010.02.001>
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435. <https://doi.org/10.2307/2183914>
- Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the Living : Episodic Memory Is Tuned to Animacy. *Current Directions in Psychological Science*, 26(1), 22-27. <https://doi.org/10.1177/0963721416667711>
- Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive Memory : The Mnemonic Value of Animacy. *Psychological Science*, 24(10), 2099-2105. <https://doi.org/10.1177/0956797613480803>
- Needham, A. (1997). Object segregation in 8-month-old infants. *Cognition*, 62(2), 121-149. [https://doi.org/10.1016/S0010-0277\(96\)00727-5](https://doi.org/10.1016/S0010-0277(96)00727-5)
- Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, 47(2), 121-148. [https://doi.org/10.1016/0010-0277\(93\)90002-D](https://doi.org/10.1016/0010-0277(93)90002-D)
- Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, 10(1-2), 3-18. <https://doi.org/10.1002/icd.239>
- Nelson, C. A., & Salapatek, P. (1986). Electrophysiological Correlates of Infant Recognition Memory. *Child Development*, 57, 1483-1497.
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104(42), 16598-16603. <https://doi.org/10.1073/pnas.0703913104>
- Newman, C., Atkinson, J., & Braddick, O. (2001). The development of reaching and looking preferences in infants to objects of different sizes. *Developmental Psychology*, 37(4), 561-572. <https://doi.org/10.1037/0012-1649.37.4.561>
- Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, 107(40), 17140-17145. <https://doi.org/10.1073/pnas.0914056107>
- Newport, E. L. (1990). Maturation constraints on language learning. *Cognitive Science*, 14, 11-28.
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottareau, 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>

- Norcia, A. M., & Gerhard, H. E. (2015). Development of Three-Dimensional Perception in Human Infants. *Annual Review of Vision Science*, *1*(1), 569-594. <https://doi.org/10.1146/annurev-vision-082114-035835>
- Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2023). Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition. *Nature Communications*, *14*(1), 8010. <https://doi.org/10.1038/s41467-023-43146-w>
- Oakes, L. M., Hurley, K. B., Ross-Sheehy, S., & Luck, S. J. (2011). Developmental changes in infants' visual short-term memory for location. *Cognition*, *118*(3), 293-305. <https://doi.org/10.1016/j.cognition.2010.11.007>
- Oakes, L. M., Messenger, I. M., Ross-Sheehy, S., & Luck, S. J. (2009). New evidence for rapid development of colour–location binding in infants' visual short-term memory. *Visual Cognition*, *17*(1-2), 67-82. <https://doi.org/10.1080/13506280802151480>
- Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior and Development*, *19*(4), 425-440. [https://doi.org/10.1016/S0163-6383\(96\)90004-1](https://doi.org/10.1016/S0163-6383(96)90004-1)
- Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid Development of Feature Binding in Visual Short-Term Memory. *Psychological Science*, *17*(9), 781-787. <https://doi.org/10.1111/j.1467-9280.2006.01782.x>
- Olejniczak, P. (2006). Neurophysiologic Basis of EEG: *Journal of Clinical Neurophysiology*, *23*(3), 186-189. <https://doi.org/10.1097/01.wnp.0000220079.61973.6c>
- Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data : Maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123-135. <https://doi.org/10.1038/nrn2314>
- Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors determining where category-selective areas emerge in visual cortex. *Trends in Cognitive Sciences*, *23*(9), 784-797. <https://doi.org/10.1016/j.tics.2019.06.006>
- Papeo, L., Wurm, M. F., Oosterhof, N. N., & Caramazza, A. (2017). The neural representation of human versus nonhuman bipeds and quadrupeds. *Scientific Reports*, *7*(1), 1-8. <https://doi.org/10.1038/s41598-017-14424-7>
- Pauen, S. (2000). Early Differentiation within the Animate Domain : Are Humans Something Special? *Journal of Experimental Child Psychology*, *75*(2), 134-151. <https://doi.org/10.1006/jecp.1999.2530>
- Pauen, S. (2002). The global-to-basic level shift in infants' categorical thinking : First evidence from a longitudinal study. *International Journal of Behavioral Development*, *26*(6), 492-499. <https://doi.org/10.1080/01650250143000445>
- Pauen, S., & Peykarjou, S. (2021). Neural correlates of human categorization in early infancy. In *The Oxford Handbook of Developmental Cognitive Neuroscience* (K.C. Cohen Kadosh). <https://doi.org/10.1093/oxfordhb/9780198827474.013.9>
- Pellicano, E., & Burr, D. (2012). When the world becomes 'too real' : A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504-510. <https://doi.org/10.1016/j.tics.2012.08.009>

- Peña, M., Arias, D., & Dehaene-Lambertz, G. (2014). Gaze Following Is Accelerated in Healthy Preterm Infants. *Psychological Science*, 25(10), 1884-1892. <https://doi.org/10.1177/0956797614544307>
- Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences*, 107(8), 3823-3828. <https://doi.org/10.1073/pnas.0914326107>
- Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier Speech Exposure Does Not Accelerate Speech Acquisition. *The Journal of Neuroscience*, 32(33), 11159-11163. <https://doi.org/10.1523/JNEUROSCI.6516-11.2012>
- Peykarjou, S. (2022). Frequency tagging with infants : The visual oddball paradigm. *Frontiers in Psychology*, 13, 1015611. <https://doi.org/10.3389/fpsyg.2022.1015611>
- Peykarjou, S., & Hoehl, S. (2013). Three-Month-Olds' Brain Responses to Upright and Inverted Faces and Cars. *Developmental Neuropsychology*, 38(4), 272-280. <https://doi.org/10.1080/87565641.2013.786719>
- Peykarjou, S., Hoehl, S., & Pauen, S. (2023). The development of visual categorization based on high-level cues. *Child Development*, 95(2), e122-e138. <https://doi.org/10.1111/cdev.14015>
- Peykarjou, S., Hoehl, S., & Pauen, S. (2024). The Development of Global-Level Categorization : Frequency Tagging EEG Responses. *Brain Sciences*, 14(6), 541. <https://doi.org/10.3390/brainsci14060541>
- Peykarjou, S., Hoehl, S., Pauen, S., & Rossion, B. (2017). Rapid Categorization of Human and Ape Faces in 9-Month-Old Infants Revealed by Fast Periodic Visual Stimulation. *Scientific Reports*, 7(1), Article 1. <https://doi.org/10.1038/s41598-017-12760-2>
- Peykarjou, S., Pauen, S., & Hoehl, S. (2014). How do 9-month-old infants categorize human and ape faces? A rapid repetition ERP study: Infants' categorization of human and ape faces. *Psychophysiology*, 51(9), 866-878. <https://doi.org/10.1111/psyp.12238>
- Pfister, H., Kaynig, V., Botha, C. P., Bruckner, S., Dercksen, V. J., Hege, H.-C., & Roerdink, J. B. T. M. (2014). Visualization in Connectomics. In *Scientific Visualization : Uncertainty, Multifield, Biomedical, and Scalable Visualization* (p. 221-245). <http://arxiv.org/abs/1206.1428>
- Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex : A functional MRI study. *Proceedings of the National Academy of Sciences*, 102(19), 6996-7001. <https://doi.org/10.1073/pnas.0502605102>
- Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Current Biology*, 19(4), 319-324. <https://doi.org/10.1016/j.cub.2009.01.007>
- Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI Reveal Dissociable Cortical Pathways for Dynamic and Static Face Perception. *Current Biology*, 24(17), 2066-2070. <https://doi.org/10.1016/j.cub.2014.07.060>
- Poulin-Dubois, D., Crivello, C., & Wright, K. (2015). Biological Motion Primes the Animate/Inanimate Distinction in Infancy. *PLOS ONE*, 10(2), 1-12. <https://doi.org/10.1371/journal.pone.0116910>
- Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, 110(41), E3965-E3972. <https://doi.org/10.1073/pnas.1304326110>

- Powell, L. J., & Spelke, E. S. (2018). Human infants' understanding of social imitation : Inferences of affiliation from third party observations. *Cognition*, *170*, 31-48. <https://doi.org/10.1016/j.cognition.2017.09.007>
- Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex : The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience*, *28*(5), 680-692.
- Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*, *193*, 167-177. <https://doi.org/10.1016/j.neuroimage.2019.03.028>
- Pyles, J. A., Verstynen, T. D., Schneider, W., & Tarr, M. J. (2013). Explicating the Face Perception Network with White Matter Connectivity. *PLoS ONE*, *8*(4), e61611. <https://doi.org/10.1371/journal.pone.0061611>
- Quek, G. L., & De Heering, A. (2024). Visual periodicity reveals distinct attentional signatures for face and non-face categories. *Cerebral Cortex*, *34*(6), bhae228. <https://doi.org/10.1093/cercor/bhae228>
- Quek, G. L., & Rossion, B. (2017). Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. *Neuropsychologia*, *104*, 182-200. <https://doi.org/10.1016/j.neuropsychologia.2017.08.010>
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, *27*(2), 145-179. [https://doi.org/10.1016/0010-0277\(87\)90017-5](https://doi.org/10.1016/0010-0277(87)90017-5)
- Quinn, P. C. (2004). Development of Subordinate-Level Categorization in 3- to 7-Month-Old Infants. *Child Development*, *75*(3), 886-899.
- Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, *63*(1), 189-211. <https://doi.org/10.1006/jecp.1996.0047>
- Quinn, P. C., & Eimas, P. D. (1998). Evidence for a global categorical representation of Humans by young infants. *Journal of Experimental Child Psychology*, *69*(3), 151-174. <https://doi.org/10.1006/jecp.1998.2443>
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463-475. <https://doi.org/10.1068/p220463>
- Quinn, P. C., Eimas, P. D., & Tarr, M. J. (2001). Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *Journal of Experimental Child Psychology*, *79*(1), 78-94. <https://doi.org/10.1006/jecp.2000.2609>
- Quinn, P. C., & Johnson, M. H. (2000). Global-Before-Basic Object Categorization in Connectionist Networks and 2-Month-Old Infants. *Infancy*, *1*(1), 31-46. https://doi.org/10.1207/S15327078IN0101_04
- Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural Markers of Categorization in 6-Month-Old Infants. *Psychological Science*, *17*(1), 59-66. <https://doi.org/10.1111/j.1467-9280.2005.01665.x>
- Rager, G., & Singer, W. (1998). The response of cat visual cortex to flicker stimuli of variable frequency. *European Journal of Neuroscience*, *10*(5), 1856-1877. <https://doi.org/10.1046/j.1460-9568.1998.00197.x>

- Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological bulletin*, *127*(2), 209-228.
- Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology*, *16*(4), 408-414. <https://doi.org/10.1016/j.conb.2006.06.004>
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, *20*(3), 238-248. [https://doi.org/10.1016/0013-4694\(66\)90088-5](https://doi.org/10.1016/0013-4694(66)90088-5)
- Regan, D. (2009). Some early uses of evoked brain responses in investigations of human visual function. *Vision Research*, *49*(9), 882-897. <https://doi.org/10.1016/j.visres.2008.01.017>
- Rekow, D., Baudouin, J.-Y., Brochard, R., Rossion, B., & Leleu, A. (2022). Rapid neural categorization of facelike objects predicts the perceptual awareness of a face (face pareidolia). *Cognition*, *222*, 105016. <https://doi.org/10.1016/j.cognition.2022.105016>
- Rekow, D., Baudouin, J.-Y., Durand, K., & Leleu, A. (2022). Smell what you hardly see : Odors assist visual categorization in the human brain. *NeuroImage*, 119181. <https://doi.org/10.1016/j.neuroimage.2022.119181>
- Rekow, D., Baudouin, J.-Y., Poncet, F., Damon, F., Durand, K., Schaal, B., Rossion, B., & Leleu, A. (2021). Odor-driven face-like categorization in the human infant brain. *Proceedings of the National Academy of Sciences*, *118*(21), e2014979118. <https://doi.org/10.1073/pnas.2014979118>
- Rekow, D., Leleu, A., Poncet, F., Damon, F., Rossion, B., Durand, K., Schaal, B., & Baudouin, J.-Y. (2020). Categorization of objects and faces in the infant brain and its sensitivity to maternal odor : Further evidence for the role of intersensory congruency in perceptual development. *Cognitive Development*, *55*, 100930. <https://doi.org/10.1016/j.cogdev.2020.100930>
- Retter, T. L., Jiang, F., Webster, M. A., Michel, C., Schiltz, C., & Rossion, B. (2021). Varying Stimulus Duration Reveals Consistent Neural Activity and Behavior for Human Face Individuation. *Neuroscience*, *472*, 138-156. <https://doi.org/10.1016/j.neuroscience.2021.07.025>
- Retter, T. L., Jiang, F., Webster, M. A., & Rossion, B. (2020). All-or-none face categorization in the human brain. *NeuroImage*, *213*, 116685. <https://doi.org/10.1016/j.neuroimage.2020.116685>
- Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, *91*, 9-28. <https://doi.org/10.1016/j.neuropsychologia.2016.07.028>
- Reynolds, G. D., & Richards, J. E. (2005). Familiarization, Attention, and Recognition Memory in Infancy : An Event-Related Potential and Cortical Source Localization Study. *Developmental Psychology*, *41*(4), 598-615. <https://doi.org/10.1037/0012-1649.41.4.598>
- Riggins, T., & Scott, L. S. (2020). P300 development from infancy to adolescence. *Psychophysiology*, *57*(7). <https://doi.org/10.1111/psyp.13346>
- Ritchie, J. B., Zeman, A. A., Bosmans, J., Sun, S., Verhaegen, K., & Op de Beeck, H. P. (2020). Untangling the animacy organization of occipitotemporal cortex. *bioRxiv*. <https://doi.org/10.1101/2020.07.17.206896>
- Ritchie, J. B., Zeman, A. A., Bosmans, J., Sun, S., Verhaegen, K., & Op de Beeck, H. P. (2021). Untangling the Animacy Organization of Occipitotemporal Cortex. *The Journal of Neuroscience*, *41*(33), 7103-7119. <https://doi.org/10.1523/JNEUROSCI.2628-20.2021>

- Rogers, T. T., & Patterson, K. (2007). Object categorization : Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451-469. <https://doi.org/10.1037/0096-3445.136.3.451>
- Rosch, E. (1978). Principles of categorization. In *Cognition and Categorization* (p. 27-48). Routledge.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances : Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.
- Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory : Implications for information processing. *Developmental Psychology*, *18*(5), 704-713. <https://doi.org/10.1037/0012-1649.18.5.704>
- Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, *18*(11), 1. <https://doi.org/10.1167/18.11.1>
- Rossion, B. (2014a). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, *18*(6), 310-318. <https://doi.org/10.1016/j.tics.2014.02.013>
- Rossion, B. (2014b). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental Brain Research*, *232*(6), 1599-1621. <https://doi.org/10.1007/s00221-014-3934-9>
- Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, *11*(2), 1-21. <https://doi.org/10.1167/11.2.16>
- Rossion, B., Retter, T. L., & Liu-Shuang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *European Journal of Neuroscience*, *52*(10), 4283-4344. <https://doi.org/10.1111/ejn.14865>
- Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*(1), 18-18. <https://doi.org/10.1167/15.1.18>
- Salapatek, P., Bechtold, A. G., & Bushnell, E. W. (1976). Infant Visual Acuity as a Function of Viewing Distance. *Child Development*, *47*, 860-863.
- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents : Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995-1001. <https://doi.org/10.1111/j.1467-9280.2005.01649.x>
- Schmidt, F., Hegele, M., & Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, *17*(11), 10. <https://doi.org/10.1167/17.11.10>
- Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences*, *110*(40), 15937-15942. <https://doi.org/10.1073/pnas.1314075110>
- Shimojo, S., Birch, E., Gwiazda, J., & Held, R. (1984). Development of vernier acuity in infants. *Vision Research*, *24*(7), 721-728.
- Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and Humans. *Journal of Cognitive Neuroscience*, *14*(2), 187-198. <https://doi.org/10.1162/089892902317236830>
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318-320. <https://doi.org/10.1038/415318a>

- Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809-813. <https://doi.org/10.1073/pnas.0707021105>
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Skelton, A. E., Maule, J., & Franklin, A. (2022). Infant color perception : Insight into perceptual development. *Child Development Perspectives*, *16*(2), 90-95. <https://doi.org/10.1111/cdep.12447>
- Skoczenski, A. M., & Norcia, A. M. (2002). Late Maturation of Visual Hyperacuity. *Psychological Science*, *13*(6), 537-541. <https://doi.org/10.1111/1467-9280.00494>
- Slater, A. (2002). Visual perception in the newborn infant : Issues and debates. *Intellectica. Revue de l'Association Pour La Recherche Cognitive*, *34*(1), 57-76. <https://doi.org/10.3406/intel.2002.1073>
- Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth : Newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology*, *49*(2), 314-322. [https://doi.org/10.1016/0022-0965\(90\)90061-c](https://doi.org/10.1016/0022-0965(90)90061-c)
- Slater, A., Morison, V., & Somers, M. (1988). Orientation Discrimination and Cortical Function in the Human Newborn. *Perception*, *17*(5), 597-602. <https://doi.org/10.1068/p170597>
- Slaughter, V., Heron, M., & Sim, S. (2002). Development of preferences for the human body shape in infancy. *Cognition*, *85*(3), B71-B81. [https://doi.org/10.1016/S0010-0277\(02\)00111-7](https://doi.org/10.1016/S0010-0277(02)00111-7)
- Soska, K. C., & Adolph, K. E. (2014). Postural Position Constrains Multimodal Object Exploration in Infants. *Infancy*, *19*(2), 138-161. <https://doi.org/10.1111/infa.12039>
- Spelke, E. S. (1994). Initial knowledge : Six suggestions. *Cognition*, *50*, 431-445.
- Spelke, E. S. (2022). *What Babies Know : Core Knowledge and Composition Volume 1*. Oxford University Press.
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of Knowledge. *Psychological Review*, *99*(4), 605-632.
- Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*(2), 113-142. <https://doi.org/10.1111/j.2044-835X.1995.tb00669.x>
- Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In *Causal cognition : A multidisciplinary debate* (p. 44-78). Clarendon Press/Oxford University Press.
- Spencer, J., Quinn, P. C., Johnson, M. H., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose : Evidence for young infants categorizing mammals by head and facial attributes. *Early Development and Parenting*, *6*(3-4), 113-126. [https://doi.org/10.1002/\(SICI\)1099-0917\(199709/12\)6:3/4<113::AID-EDP151>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<113::AID-EDP151>3.0.CO;2-W)
- Spriet, C., Abassi, E., Hochmann, J.-R., & Papeo, L. (2022). Visual object categorization in infancy. *Proceedings of the National Academy of Sciences*, *119*(8), e2105866119. <https://doi.org/10.1073/pnas.2105866119>

- Storrs, K. R., Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2020). Noise ceiling on the crossvalidated performance of reweighted models of representational dissimilarity: Addendum to Khaligh-Razavi & Kriegeskorte (2014). *bioRxiv*. <https://doi.org/10.1101/2020.03.23.003046>
- Stothart, G., Quadflieg, S., & Milton, A. (2017). A fast and implicit measure of semantic categorisation using steady state visual evoked potentials. *Neuropsychologia*, *102*, 11-18. <https://doi.org/10.1016/j.neuropsychologia.2017.05.025>
- Striano, T., & Bertin, E. (2005). Social-cognitive skills between 5 and 10 months of age. *British Journal of Developmental Psychology*, *23*(4), 559-568. <https://doi.org/10.1348/026151005X26282>
- Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds. *Current Biology*, *24*(6), 687-692. <https://doi.org/10.1016/j.cub.2014.02.010>
- Striem-Amit, E., Dakwar, O., Reich, L., & Amedi, A. (2012). The large-scale organization of “visual” streams emerges without visual experience. *Cerebral Cortex*, *22*(7), 1698-1709. <https://doi.org/10.1093/cercor/bhr253>
- Surian, L., & Caldi, S. (2010). Infants’ individuation of agents and inert objects. *Developmental Science*, *13*(1), 143-150. <https://doi.org/10.1111/j.1467-7687.2009.00873.x>
- Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Teller, D. Y., Peeples, D. R., & Sekel, M. (1978). Discrimination of chromatic from white light by two-month-old human infants. *Vision Research*, *18*(1), 41-48. [https://doi.org/10.1016/0042-6989\(78\)90075-5](https://doi.org/10.1016/0042-6989(78)90075-5)
- Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *eLife*, *8*:e47142. <https://doi.org/10.7554/eLife.47142>
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522. <https://doi.org/10.1038/381520a0>
- Tiedemann, H., Schmidt, F., & Fleming, R. W. (2022). Superordinate Categorization Based on the Perceptual Organization of Parts. *Brain Sciences*, *12*(5), 667. <https://doi.org/10.3390/brainsci12050667>
- Tompa, T., & Sáry, G. (2010). A review on the inferior temporal cortex of the macaque. *Brain Research Reviews*, *62*(2), 165-182. <https://doi.org/10.1016/j.brainresrev.2009.10.001>
- Tsuneishi, S., & Casaer, P. (2007). Effects of preterm extrauterine visual experience on the development of the human visual system: A flash VEP study. *Developmental Medicine & Child Neurology*, *42*(10), 663-668. <https://doi.org/10.1111/j.1469-8749.2000.tb00676.x>
- Tyler, L., & Moss, H. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*(6), 244-252. [https://doi.org/10.1016/S1364-6613\(00\)01651-X](https://doi.org/10.1016/S1364-6613(00)01651-X)
- Ungerleider, L. G., & Haxby, J. V. (1994). « What » and « where » in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157-165. [https://doi.org/10.1016/0959-4388\(94\)90066-3](https://doi.org/10.1016/0959-4388(94)90066-3)
- Valentini, N. C., Pereira, K. R. G., Chiquetti, E. M. D. S., Formiga, C. K. M. R., & Linhares, M. B. M. (2019). Motor trajectories of preterm and full-term infants in the first year of life. *Pediatrics International*, *61*(10), 967-977. <https://doi.org/10.1111/ped.13963>

- Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 892-903. <https://doi.org/10.1037/0096-1523.22.4.892>
- Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology*, 3(7), e208. <https://doi.org/10.1371/journal.pbio.0030208>
- Van Blooijis, D., Van Den Boom, M. A., Van Der Aar, J. F., Huiskamp, G. M., Castegnaro, G., Demuru, M., Zweiphenning, W. J. E. M., Van Eijsden, P., Miller, K. J., Leijten, F. S. S., & Hermes, D. (2023). Developmental trajectory of transmission speed in the human brain. *Nature Neuroscience*, 26(4), 537-541. <https://doi.org/10.1038/s41593-023-01272-0>
- Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds : Predictive coding in autism. *Psychological Review*, 121(4), 649-675. <https://doi.org/10.1037/a0037665>
- van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, 114(22), E4501-E4510. <https://doi.org/10.1073/pnas.1612862114>
- Van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 Development across the Lifespan : A Systematic Review and Meta-Analysis. *PLoS ONE*, 9(2), e87347. <https://doi.org/10.1371/journal.pone.0087347>
- Van Hof-Van Duin, J., & Mohn, G. (1986). The development of visual acuity in normal fullterm and preterm infants. *Vision Research*, 26(6), 909-916. [https://doi.org/10.1016/0042-6989\(86\)90149-5](https://doi.org/10.1016/0042-6989(86)90149-5)
- Vanmarcke, S., van Esch, L., Van der Hallen, R., Evers, K., Noens, I., Steyaert, J., & Wagemans, J. (2016). Gist perception in adolescents with and without ASD : Ultra-rapid categorization of meaningful real-life scenes. *Research in Autism Spectrum Disorders*, 29-30, 30-47. <https://doi.org/10.1016/j.rasd.2016.05.007>
- VanRullen, R., & Thorpe, S. J. (2001). Is it a Bird? Is it a Plane? Ultra-Rapid Visual Categorisation of Natural and Artificial Objects. *Perception*, 30(6), 655-668. <https://doi.org/10.1068/p3029>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172-1217. <https://doi.org/10.1037/a0029333>
- Wandell, B. A. (1995). The Photoreceptor Mosaic. In *Foundations of vision (chapitre 3)* (p. xvi, 476). Sinauer Associates.
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, 56(2), 366-383. <https://doi.org/10.1016/j.neuron.2007.10.012>
- Wang, R., Janini, D., & Konkle, T. (2022). Mid-level Feature Differences Support Early Animacy and Object Size Distinctions : Evidence from Electroencephalography Decoding. *Journal of Cognitive Neuroscience*, 34(9), 1670-1680. https://doi.org/10.1162/jocn_a_01883
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829-854.
- Watanabe, M., & Rodieck, R. W. (1989). Parasol and midget ganglion cells of the primate retina. *Journal of Comparative Neurology*, 289(3), 434-454. <https://doi.org/10.1002/cne.902890308>

- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories : Evidence from 12- to 13-month-old infants. *Cognitive psychology*, *29*(3), 257-302.
- Wertz, A. E., & Wynn, K. (2014a). Selective Social Learning of Plant Edibility in 6- and 18-Month-Old Infants. *Psychological Science*, *25*(4), 874-882. <https://doi.org/10.1177/0956797613516145>
- Wertz, A. E., & Wynn, K. (2014b). Thyme to touch : Infants possess strategies that protect them from dangers posed by plants. *Cognition*, *130*(1), 44-49. <https://doi.org/10.1016/j.cognition.2013.09.002>
- Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120391. <https://doi.org/10.1098/rstb.2012.0391>
- Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex : Evidence for task-independent category effects. *Neuropsychologia*, *47*(14), 3111-3117. <https://doi.org/10.1016/j.neuropsychologia.2009.07.008>
- Wilcox, T. (1999). Object individuation : Infants' use of shape, size, pattern, and color. *Cognition*, *72*, 125-166.
- Wilcox, T., Woods, R., & Chapa, C. (2008). Color–function categories that prime infants to use color information in an object individuation task. *Cognitive Psychology*, *57*(3), 220-261. <https://doi.org/10.1016/j.cogpsych.2008.02.001>
- Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34. [https://doi.org/10.1016/S0010-0277\(98\)00058-4](https://doi.org/10.1016/S0010-0277(98)00058-4)
- Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, *22*(2), 145-160. [https://doi.org/10.1016/S0163-6383\(99\)00007-7](https://doi.org/10.1016/S0163-6383(99)00007-7)
- Xie, S., Hoehl, S., Moeskops, M., Kayhan, E., Kliesch, C., Turtleton, B., Köster, M., & Cichy, R. M. (2022). Visual category representations in the infant brain. *Current Biology*, *32*(24), 5422-5432.e6. <https://doi.org/10.1016/j.cub.2022.11.016>
- Xu, F., & Carey, S. (1996). Infants' metaphysics : The case of numerical identity. *Cognitive Psychology*, *30*(2), 111-153. <https://doi.org/10.1006/cogp.1996.0005>
- Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology*, *49*(2), 155-190. <https://doi.org/10.1016/j.cogpsych.2004.01.001>
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619-8624. <https://doi.org/10.1073/pnas.1403112111>
- Yan, X., Tung, S., Fascendini, B., Chen, Y. D., Norcia, A. M., & Grill-Spector, K. (2023). When Do Visual Category Representations Emerge in Infants' Brains? *bioRxiv*. <https://doi.org/10.1101/2023.05.11.539934>
- Zemach, I., Chang, S., & Teller, D. Y. (2007). Infant color vision : Prediction of infants' spontaneous color preferences. *Vision Research*, *47*(10), 1368-1381. <https://doi.org/10.1016/j.visres.2006.09.024>

- Zeman, A. A., Ritchie, J. B., Bracci, S., & Op de Beeck, H. (2020). Orthogonal Representations of Object Shape and Category in Deep Convolutional Neural Networks and Human Visual Cortex. *Scientific Reports*, *10*(1), 2453. <https://doi.org/10.1038/s41598-020-59175-0>
- Zhang, H., Tian, J., Liu, J., Li, J., & Lee, K. (2009). Intrinsically organized network for face perception during the resting state. *Neuroscience Letters*, *454*(1), 1-5. <https://doi.org/10.1016/j.neulet.2009.02.054>
- Zieber, N., Bhatt, R. S., Hayden, A., Kangas, A., Collins, R., & Bada, H. (2010). Body representation in the first year of life. *Infancy*, *15*(5), 534-544. <https://doi.org/10.1111/j.1532-7078.2009.00026.x>

9. Supplementary materials

9.1. Chapter 3 – Supplementary results

1. Stepwise linear regression analysis (Experiments 1-2)

We run a stepwise linear regression with the six synthetic models as regressors for each group in Experiments 1-2. As reported in the main text, this analysis showed an effect of the eight-category model, animacy model, and humanness model for 19-month-old infants (Experiment 1), and of the animacy model only for 10-month-old infants (Experiment 1) and 4-month-old infants in Experiment 2. See Supplementary Table 1 for statistical values of this analysis.

Supplementary Table 1. Results of representational similarity analysis (Stepwise linear regression) reflecting relationships between the infants' DLT-RDMs and the synthetic models of categorization

Exp.	Age	Regressor	Mean β (<i>SD</i>)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1	19 m	Eight-categories	.090 (.112)	.026 – .155	4.023 (24)	<.001	.804
		Animacy	.077 (.109)	.014 – .139	3.514 (24)	.002	.702
		Humanness	.133 (.215)	.009 – .256	3.091 (24)	.005	.618
		Faces/Bodies	.022 (.135)	-.056 – .099	0.811 (24)	<i>n.s.</i>	.163
		Natural/Artificial	.017 (.219)	-.109 – .143	0.384 (24)	<i>n.s.</i>	.078
		Big/Small	.066 (.176)	-.036 – .167	1.866 (24)	.074	.375
1	10 m	Eight-categories	.051 (.113)	-.015 – .118	2.222 (23)	.036	.451
		Animacy	.059 (.098)	.002 – .117	2.981 (23)	.007	.608
		Humanness	.002 (.200)	-.116 – .119	0.039 (23)	<i>n.s.</i>	.010
		Faces/Bodies	.045 (.181)	-.062 – .152	1.207 (23)	.240	.249
		Natural/Artificial	-.026 (.122)	-.098 – .046	-1.054 (23)	<i>n.s.</i>	.213
		Big/Small	.038 (.128)	-.038 – .113	1.448 (23)	.161	.297
1	4 m	Eight-categories	.013 (.101)	-.047 – .073	0.634 (23)	<i>n.s.</i>	.129
		Animacy	.014 (.118)	-.055 – .084	0.598 (23)	<i>n.s.</i>	.119
		Humanness	.034 (.206)	-.088 – .156	0.809 (23)	<i>n.s.</i>	.165
		Faces/Bodies	.012 (.200)	-.106 – .130	0.304 (23)	<i>n.s.</i>	.060
		Natural/Artificial	.031 (.250)	-.117 – .178	0.606 (23)	<i>n.s.</i>	.124
		Big/Small	.021 (.246)	-.124 – .166	0.421 (23)	<i>n.s.</i>	.085
2	4 m	Eight-categories	.026 (.142)	-.057 – .110	0.906 (23)	<i>n.s.</i>	.183
		Animacy	.074 (.098)	.016 – .132	3.697 (23)	.001	.755
		Humanness	-.061 (.223)	-.192 – .070	-1.341 (23)	.193	.274
		Faces/Bodies	.019 (.211)	-.106 – .143	0.430 (23)	<i>n.s.</i>	.090
		Natural/Artificial	.096 (.296)	-.079 – .271	1.591 (23)	.125	.324
		Big/Small	.008 (.232)	-.129 – .145	0.165 (23)	<i>n.s.</i>	.034

Note: Exp., experiment; m, months; CI, 99.17% confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, two-tailed; *n.s.* = non-significant results with $P > .250$.

2. Difference between within-category and between-category DLTs (Experiments 1-2).

For each Experiment, for each age group, for each of the six categorization models, we tested whether within-category differential looking times (DLTs) were lower than between-category DLTs. As mentioned in the main text, this analysis showed an effect of the eight-category model, animacy model, and humanness model in 19-month-old infants (Experiment 1), and of the animacy model only in 10-month-old infants (Experiment 1) and in 4-month-old infants of Experiment 2. Supplementary Table 2 report all statistical values of this analysis.

Supplementary Table 2. Results of the DLTs analyses of within versus between category comparisons

Exp.	Age	Comparisons within/between	Mean of the difference (SD)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1	19 m	Diagonal	-.135 (.085)	-inf – -.091	-7.967 (24)	<.0001	1.588
		Animate and inanimate	-.095 (.106)	- inf – -.041	-4.481 (24)	<.0001	.896
		Human and nonhuman	-.158 (.190)	- inf – -.060	-4.160 (24)	<.001	.832
		Faces and Bodies	-.057 (.141)	- inf – .016	-2.013 (24)	.028	.404
		Natural and Artificial	-.035 (.210)	- inf – .074	-.825 (24)	.209	.167
		Big and Small	-.074 (.178)	- inf – .018	-2.071 (24)	.025	.416
1	10 m	Diagonal	-.057 (.137)	- inf – .016	-2.018 (23)	.028	.416
		Animate and inanimate	-.061 (.103)	- inf – -.007	-2.919 (23)	.004	.592
		Human and nonhuman	-.003 (.202)	- inf – .104	-.066 (23)	<i>n.s.</i>	.015
		Faces and Bodies	-.045 (.202)	- inf – .061	-1.096 (23)	.142	.223
		Natural and Artificial	.029 (.139)	-inf – .102	1.036 (23)	<i>n.s.</i>	.209
		Big and Small	-.041 (.145)	- inf – .036	-1.370 (23)	.092	.283
1	4 m	Diagonal	-.048 (.120)	- inf – .016	-1.940 (23)	.032	.400
		Animate and inanimate	-.022 (.121)	- inf – .042	-.873 (23)	.196	.182
		Human and nonhuman	-.036 (.216)	-inf – .078	-.811 (23)	.213	.167
		Faces and Bodies	-.013 (.205)	- inf – .095	-.316 (23)	<i>n.s.</i>	.063
		Natural and Artificial	-.043 (.252)	- inf – .090	-.840 (23)	.205	.171
		Big and Small	-.027 (.262)	- inf – .111	-.513 (23)	<i>n.s.</i>	.103
2	4 m	Diagonal	-.024 (.133)	- inf – .046	-.891 (23)	.191	.180
		Animate and inanimate	-.076 (.104)	- inf – -.021	-3.583 (23)	<.001	.731
		Human and nonhuman	.087 (.232)	- inf – .209	1.836 (23)	<i>n.s.</i>	.375
		Faces and Bodies	.018 (.225)	- inf – .136	.383 (23)	<i>n.s.</i>	.080
		Natural and Artificial	-.086 (.308)	- inf – .077	-1.360 (23)	.094	.279
		Big and Small	-.014 (.259)	- inf – .123	-.267 (23)	<i>n.s.</i>	.054

Note: Exp., experiment; m, months; CI, 99.17% confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, one-tailed; *n.s.* = non-significant results with $P > .250$.

3. Analysis of mean looking times (Experiments 1-2)

Separate one-way ANOVAs for 4-, 10- and 19-month-olds in Experiment 1 and for 4-month-olds in Experiment 2, revealed the effect of categories on mean looking times (MLTs). The following post-hoc analyses were carried out to follow up on the effect of category. For 19-month-old infants in Experiment 1, pairwise comparisons (one sample t -tests, $\alpha_{\text{corrected}}$: 0.0018) revealed longer looking times towards animate than inanimate objects (9 out of 16 comparisons were significant; 4 of the 7 remaining comparisons showed non-significant trends, that is, effects that did not reach the significance level, corrected for multiple comparisons). A t -test comparing all animate *vs.* all inanimate categories confirmed this effect (animate: $M \pm SD = 2.103 \text{ s} \pm 0.344$; inanimate: $M \pm SD = 1.598 \text{ s} \pm 0.247$; $M_{\text{difference}} \pm SD = 0.505 \pm 0.335$; 95% CI = 0.367 – 0.643; $t(24) = 7.551$, $P < 0.0001$; $d = 1.509$). Comparisons between human and nonhuman categories showed significant or non-significant trends (P s < .03), which can be summarized as a preference for nonhuman categories ($M \pm SD = 2.317 \text{ s} \pm 0.463$) over human ($M \pm SD = 1.889 \text{ s} \pm 0.452$; $M_{\text{difference}} \pm SD = -0.428 \pm 0.602$; 95% CI = -0.676 – -0.179; $t(24) = -3.552$, $P = 0.002$; $d = 0.711$). Among the inanimate categories, only the comparison between natural-small *vs.* artificial-big showed a non-significant trend ($P = 0.019$; for other comparisons all P s > 0.11).

The same analyses on 10-month-old infants (Experiment 1) showed that infants looked longer at the animate than inanimate categories (10 out of 16 comparisons were significant with $\alpha_{\text{corrected}} = 0.0018$; 4 of the 6 remaining comparisons showed non-significant trends with P s < 0.05; for the remaining two comparisons, P s > 0.062). A t -test comparing MLTs for all animate *vs.* all inanimate objects showed significantly longer MLTs for animate ($M \pm SD = 2.044 \text{ s} \pm 0.316$) than for inanimate categories ($M \pm SD = 1.601 \text{ s} \pm 0.263$; $M_{\text{difference}} \pm SD = 0.443 \pm 0.314$; 95% CI = 0.310 – 0.576; $t(23) = 6.901$, $P < 0.0001$; $d = 1.409$). When two animate or two inanimate categories were compared, only 1 out of 12 comparisons showed a non-significant trend ($P = 0.007$; for all other comparisons P s > 0.056).

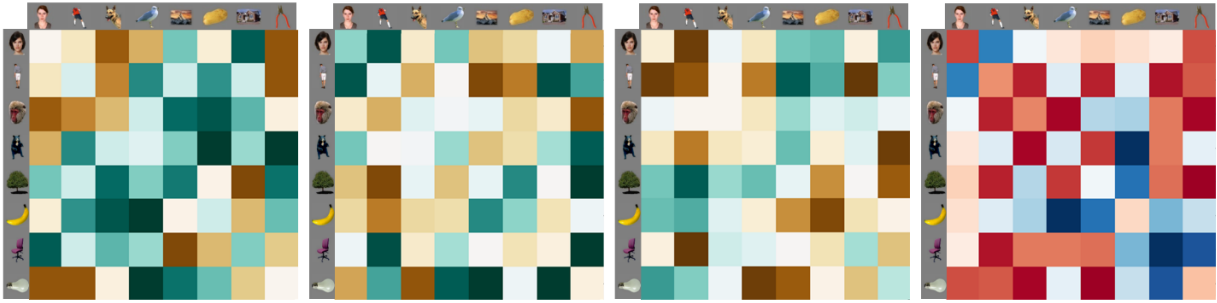
For 4-month-old infants in Experiment 1, pairwise comparisons (one sample t -tests, $\alpha_{\text{corrected}} = 0.0018$, two-tailed) revealed that infants looked longer at human faces relative to all other categories (P s < 0.001 for 6 of the 7 comparisons between the human face category and other categories respectively), except the natural big objects, for which the MLTs only marginally differed ($P = 0.002$). No other difference was observed within the animate categories. Within the inanimate categories, infants looked longer at big than small objects, whether artificial or natural (P s < 0.0001).

Four-month-old infants in Experiment 2 showed to prefer (i.e., looked longer at) human faces vs. all other categories (all P s < 0.0001) and big vs. small (natural or artificial) inanimate objects (all P s < 0.001). Unlike in Experiment 1, the preference for human faces and big objects in Experiment 2 was not conflated with the preference for large images, as all images featured the same number of pixels. However, the mean looking time for individual images, averaged across participants, negatively correlated with the image shape elongation ($\rho = -0.396$, $p < 0.001$): the less elongated the shape, the longer the looking time (see also Supplementary Results 4). See Supplementary Table 3 for t -values and P -values of all the pairwise comparisons reported here.

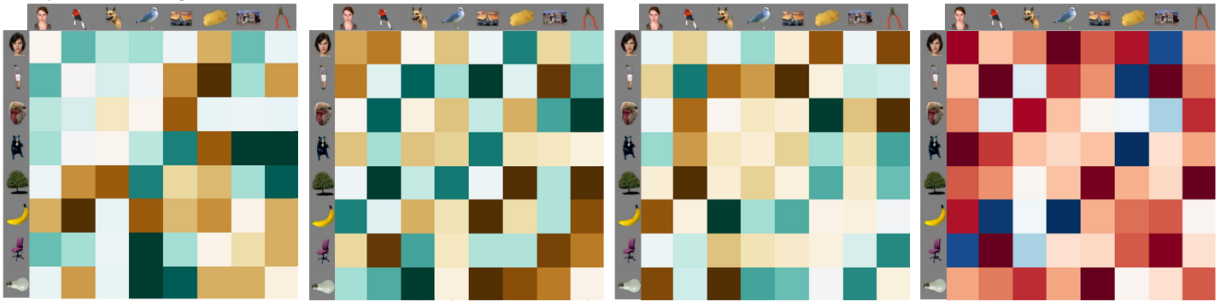
4. Relationship between MLTs and low-level features of images.

Because 4-month-olds' looking behavior appeared to be guided by low-level features such as the size of images or the elongation (Supplementary Fig. 1), we tested whether the categories of images that infants looked at for a longer time, corresponded to the categories including larger images and/or less elongated objects. To this end, we computed the MLTs, the mean number of pixels and the mean elongation ratio for each category (Supplementary Fig. 2). The three preferred categories, human faces and natural/artificial big objects, constituted the largest (Experiment 1) and least elongated (Experiments 1-2) images. Furthermore, mean looking times (MLTs) correlated significantly with both size and elongation in Experiment 1, and with elongation in Experiment 2 (see Main text).

a. Exp. 1 – Subject 2 – 19 months



b. Exp. 1 – Subject 9 – 10-months



c. Exp. 1 – Subject 2 – 4 months



d. Exp. 2 – Subject 8 – 4 months


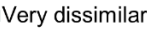




Signed matrix

Size's matrix

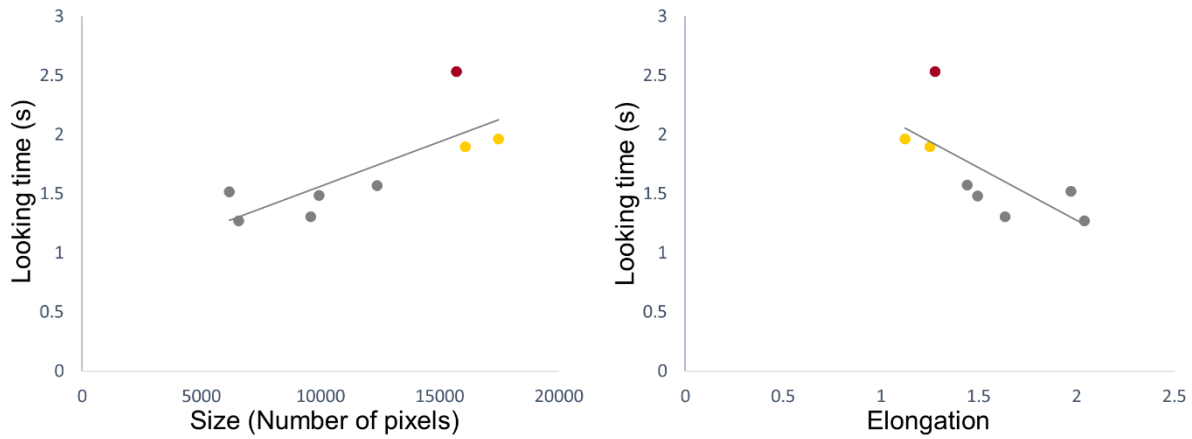
Elongation's matrix

Color's matrix

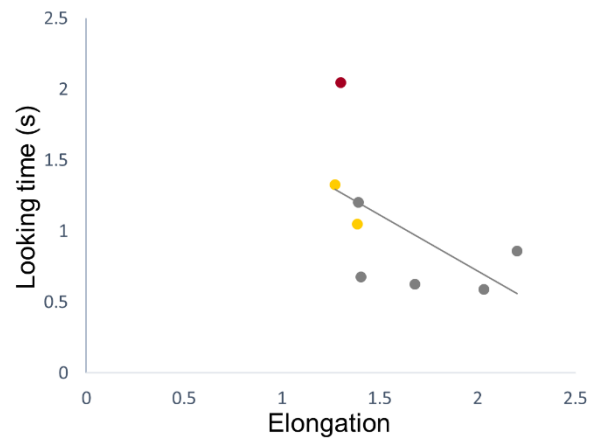
Very dissimilar  Very similar  Very dissimilar  Very similar  Very dissimilar

Supplementary Figure 1. Signed DLT-RDM (left), size RDM (middle left), elongation RDM (middle right) and color RDM (right) for 4-, 10- and 19-month-olds in Experiment 1 and 4-month-olds in Experiment 2. Example of a signed DLT-RDM, a size RDM, an elongation RDM and a color RDM for one subject of the 19-month-old group (a), the 10-month-old group (b) and the 4-month-old group (c) in Experiment 1, and one subject of the 4-month-old group in Experiment 2 (d).

a. 4-month-olds – Exp. 1



b. 4-month-olds – Exp. 2



● Human face category ● Natural and Artificial big categories

Supplementary Figure 2. Looking times as a function of image size (left) and elongation (right) for 4-month-olds in Experiments 1-2. Relationship between looking times (s) and average size of the images (number of pixels) for each category and average elongation (ratio between width and height) for each category for 4-month-olds in Experiment 1 (a) and relationship between looking times (s) and average elongation for each category for 4-month-olds in Experiment 2 (b).

Given these results, for each group of 4-month-olds, we performed a novel stepwise linear regression analysis and DLTs analysis, after removing the variance explained by size and elongation in Experiment 1, and by elongation only in Experiment 2 (Supplementary Tables 4- 5).

Supplementary Table 4. Results of the stepwise linear regression with artificial models, removing variance explained by size and/or elongation of 4-month-old infants' DLT-RDM in experiment 1 and 2

Exp.	Regressor	Mean β (<i>SD</i>)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1	Eight-categories	-.002 (.101)	-.062 – .057	-.117 (23)	<i>n.s.</i>	.020
	Animacy	.012 (.142)	-.072 – .096	.416 (23)	<i>n.s.</i>	.085
	Humanness	.045 (.246)	-.100 – .189	.887 (23)	<i>n.s.</i>	.183
	Faces/Bodies	-.026 (.237)	-.166 – .114	.531 (23)	<i>n.s.</i>	.110
	Natural/Artificial	-.029 (.186)	-.139 – .081	.755 (23)	<i>n.s.</i>	.156
	Big/Small	-.042 (.179)	-.148 – .064	1.148 (23)	<i>n.s.</i>	.235
2	Eight-categories	-.020 (.121)	-.091 – .051	-.806 (23)	<i>n.s.</i>	.165
	Animacy	.064 (.096)	.008 – .121	3.282 (23)	.003	.667
	Humanness	-.069 (.212)	-.194 – .056	-1.584 (23)	.127	.325
	Faces/Bodies	-.013 (.204)	-.133 – .108	-.303 (23)	<i>n.s.</i>	.064
	Natural/Artificial	.085 (.285)	-.084 – .253	1.453 (23)	.160	.298
	Big/Small	-.016 (.215)	-.142 – .111	-.365 (23)	<i>n.s.</i>	.074

Note: Exp., experiment; CI, confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, two-tailed; *n.s.* = non-significant results with $P > .250$.

Supplementary Table 5. Results of the DLTs analysis of within versus between category comparisons, removing variance explained by size and/or elongation of 4-month-old infants in experiment 1 and 2

Exp.	Comparisons within/between	Mean of the difference (<i>SD</i>)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1	Diagonal	-.021 (.110)	- inf – .037	-.943 (23)	.178	.191
	Animate and inanimate	-.016 (.143)	- inf – .059	-.543 (23)	<i>n.s.</i>	.112
	Human and nonhuman	-.048 (.247)	- inf – .082	-.955 (23)	.175	.194
	Faces and Bodies	.015 (.238)	- inf – .141	.310 (23)	<i>n.s.</i>	.063
	Natural and Artificial	.041 (.203)	- inf – .148	.998 (23)	<i>n.s.</i>	.202
	Big and Small	.046 (.179)	- inf – .141	1.257 (23)	<i>n.s.</i>	.257
2	Diagonal	.002 (.119)	- inf – .064	.075 (23)	<i>n.s.</i>	.017
	Animate and inanimate	-.068 (.101)	- inf – -.015	-3.314 (23)	.002	.673
	Human and nonhuman	.090 (.226)	- inf – .209	1.937 (23)	<i>n.s.</i>	.398
	Faces and Bodies	.043 (.221)	- inf – .160	.959 (23)	<i>n.s.</i>	.195
	Natural and Artificial	-.079 (.307)	- inf – .083	-1.256 (23)	.111	.257
	Big and Small	.018 (.248)	- inf – .149	.355 (23)	<i>n.s.</i>	.073

Note: Exp., experiment; CI, confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, two-tailed; *n.s.* = non-significant results with $P > .250$

5. Experiment 2: Data analysis with more lenient inclusion criteria

In Experiment 2, the attrition rate for 4-month-olds was larger compared to 4-month-olds in Experiment 1 (37% in Exp. 1 and 50% in Exp. 2). Data from 24 out of 48 tested infants were discarded in Experiment 2, while data from 14 out of 38 tested infants were discarded in Experiment 1. A larger attrition rate may lead to the inclusion of more attentive infants, which could explain the different results in Experiment 2 compared to Experiment 1.

To take into account this possibility, we ran a novel analysis on the data from Experiment 2, using inclusion criteria that yielded an attrition rate comparable to that of Experiment 1. In particular, we modified the criterion for trial inclusion, selecting all trials in which infants looked at images for at least 800 ms (instead of 1000 ms in the original analysis reported in the main text). This change yielded the exclusion of 16 out of 48 infants (attrition rate of 33%). The analysis of this dataset confirmed the results reported in the main text. The stepwise linear regression analyzing the structure of the absolute RDMS identified only one significant regressor ($\alpha_{\text{corrected}} = 0.0083$), corresponding to the animacy model (mean $\beta \pm SD = 0.066 \pm 0.094$; 99.17% CI = 0.020 – 0.113; $t(31) = 3.991$; $P < 0.001$; $d = 0.705$) (for other regressors, $ts < 2.313$; $Ps > 0.028$). The DLTs analysis also showed that the difference between within-category and between-category DLTs was only significant for the animacy model ($M_{\text{difference}} \pm SD = -0.065 \pm 0.094$; 99.17% CI = -inf – -0.023; $t(31) = -3.927$; $P < 0.001$; $d = 0.691$) (for other models, $ts > -2.208$; $Ps > 0.017$).

9.2. Chapter 4 – Supplementary results

1. Stepwise linear regression analysis (Exps. 1-3)

We first performed a correlation between infants DLT-RDMs and the set of categorization models based on fMRI responses evoked in human adults with RDMs computed at different location in the ventral stream (EVC, VOTC and LOTC; see Supplementary Table 1), as well as with the composite model, revealing a correlation for 6-month-old full-term infants in Exp. 1b and for 8-month-old full-term infants in Exp. 2 (6-month-old in Exp. 1a: $M_{\text{fisher}\rho} = 0.046 \pm 0.183$ SD; 95% CI = -0.032 – 0.123; $t(23) = 1.220$; $P = 0.235$; $d = 0.251$; 6-month-old in Exp. 1b: $M_{\text{fisher}\rho} = 0.096 \pm 0.170$ SD; 95% CI = 0.024 – 0.168; $t(23) = 2.766$; $P = 0.011$; $d = 0.565$; 8-month-old in Exp. 2: $M_{\text{fisher}\rho} = 0.192 \pm 0.171$ SD; 95% CI = 0.120 – 0.265; $t(23) = 5.521$; $P < 0.0001$; $d = 1.123$; 8-month-old in Exp. 3: $M_{\text{fisher}\rho} = 0.062 \pm 0.159$ SD; 95% CI = -0.005 – 0.130; $t(23) = 1.926$; $P = 0.067$; $d = 0.390$). We run a stepwise linear regression with the six synthetic models as regressors for each group in Exps. 1-3. As reported in the main text, this analysis showed an effect of the eight-category model and the animacy model for 8-month-old infants (Exp. 2). See Supplementary Table 2 for statistical values of this analysis.

Supplementary Table 1. Results of representation similarity analysis reflecting relationships between the infants' DLT-RDMs and the synthetic models of categorization

Exp.	Age	Model	Mean ρ (SD)	CI (min – max)	t (df)	P	Cohen's d
1a	6 m	EVC	-.061 (.160)	-.128 – .007	-1.870 (23)	.074	.382
		VOTC	.024 (.152)	-.040 – .089	.781 (23)	<i>n.s.</i>	.160
		LOTC	-.014 (.189)	-.093 – .066	-.350 (23)	<i>n.s.</i>	.072
1b	6 m	EVC	-.032 (.187)	-.111 – .048	-.825 (23)	<i>n.s.</i>	.168
		VOTC	.063 (.149)	-.00004 – .126	2.067 (23)	.050	.422
		LOTC	.043 (.176)	-.032 – .117	1.188 (23)	.247	.243
2	8 m	EVC	.087 (.110)	.041 – .133	3.894 (23)	<.001	0.795
		VOTC	.171 (.131)	.116 – .227	6.381 (23)	<.0001	1.303
		LOTC	.152 (.121)	.101 – .203	6.165 (23)	<.0001	1.258
3	8 m	EVC	.004 (.216)	-.087 – .096	.100 (23)	<i>n.s.</i>	.019
		VOTC	.049 (.170)	-.023 – .121	1.408 (23)	.173	.288
		LOTC	.040 (.174)	-.033 – .114	1.136 (23)	<i>n.s.</i>	.230

Note: Exp., experiment; Age is the chronological age; m, months; EVC, early visual cortex; VOTC, ventral occipitotemporal cortex; LOTC, lateral occipitotemporal cortex; mean ρ are the fisher transformed ρ ; CI, confidence interval; Highlighted in bold are the significant results; $\alpha = .017$ for EVC, VOTC and LOTC; *n.s.* = non-significant results with $P > .250$.

Because 8-month-old infants' behavior showed evidence of categorization based on the eight-category model, we wondered which of the eight categories infants indeed represented. Thus, for each category, we tested whether the average within-category DLTs were lower than the average between-category DLTs (t -tests; $\alpha_{\text{corrected}}$: 0.0063, one-tailed). We found that, in addition to animates and inanimates, 8-month-olds represented the subordinate categories of human faces ($M_{\text{difference}} = -0.270$; 99.37% CI = -inf – -0.120; $t(23) = -4.868$, $P < 0.0001$; $d = 0.993$) and of natural small objects ($M_{\text{difference}} = -0.152$; 99.37% CI = -inf – -0.031; $t(23) = -3.399$, $P = 0.001$; $d = 0.710$) (all other P s > .03).

Supplementary Table 2. Results of representational similarity analysis (Stepwise linear regression) reflecting relationships between the infants' DLT-RDMs and the synthetic models of categorization

Exp.	Age	Regressor	Mean β (SD)	CI (min – max)	t (df)	P	Cohen's d
1a	6 m	Eight-categories	.010 (.120)	-.061 – .080	0.407 (23)	<i>n.s.</i>	.083
		Animacy	.022 (.086)	-.029 – .073	1.260 (23)	.220	.256
		Humanness	.013 (.234)	-.125 – .151	0.272 (23)	<i>n.s.</i>	.056
		Faces/Bodies	-.031 (.163)	-.127 – .065	-0.919 (23)	<i>n.s.</i>	.190
		Natural/Artificial	-.003 (.149)	-.091 – .086	-0.084 (23)	<i>n.s.</i>	.020
		Big/Small	.001 (.161)	-.094 – .096	0.033 (23)	<i>n.s.</i>	.006
1b	6 m	Eight-categories	.025 (.112)	-.041 – .091	1.090 (23)	<i>n.s.</i>	.223
		Animacy	.041 (.086)	-.009 – .092	2.350 (23)	.028	.477
		Humanness	.042 (.248)	-.105 – .188	.820 (23)	<i>n.s.</i>	.169
		Faces/Bodies	.053 (.177)	-.052 – .157	1.464 (23)	.157	.299
		Natural/Artificial	-.036 (.173)	-.138 – .066	-1.012 (23)	<i>n.s.</i>	.208
		Big/Small	.027 (.245)	-.118 – .171	.534 (23)	<i>n.s.</i>	.110
2	8 m	Eight-categories	.089 (.080)	.042 – .136	5.442 (23)	<.0001	1.113
		Animacy	.099 (.099)	.041 – .157	4.914 (23)	<.0001	1.000
		Humanness	.044 (.145)	-.041 – .129	1.500 (23)	.147	.303
		Faces/Bodies	.058 (.201)	-.060 – .176	1.413 (23)	.171	.289
		Natural/Artificial	-.020 (.159)	-.114 – .073	-.631 (23)	<i>n.s.</i>	.126
		Big/Small	.077 (.181)	-.030 – .184	2.089 (23)	.048	.425
3	8 m	Eight-categories	.036 (.101)	-.024 – .095	1.724 (23)	.098	.356
		Animacy	.050 (.110)	-.015 – .114	2.213 (23)	.037	.455
		Humanness	.062 (.169)	-.037 – .162	1.811 (23)	.083	.367
		Faces/Bodies	.083 (.221)	-.048 – .213	1.834 (23)	.080	.376
		Natural/Artificial	-.072 (.162)	-.168 – .023	-2.193 (23)	.039	.444
		Big/Small	-.035 (.176)	-.138 – .069	-.966 (23)	<i>n.s.</i>	.199

Note: Exp., experiment; m, months; CA, corrected age; CI, 99.17% confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, two-tailed; *n.s.* = non-significant results with $P > .250$.

2. Difference between within-category and between-category DLTs (Exps. 1-3).

For each Experiment, for each age group, for each of the six categorization models, we tested whether within-category differential looking times (DLTs) were lower than between-category DLTs. As

mentioned in the main text, this analysis showed an effect of the eight-category model, animacy model, and big-small real-world size model in 8-month-old infants (Exp. 2). Supplementary Table 3 report all statistical values of this analysis.

Supplementary Table 3. Results of the DLTs analyses of within versus between category comparisons

Exp.	Age	Comparisons within/between	Mean of the difference (<i>SD</i>)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1a	6 m	Eight-categories	-.006 (.119)	- inf – .056	-.257 (23)	<i>n.s.</i>	.050
		Animacy	-.024 (.092)	- inf – .024	-1.279 (23)	.107	.261
		Human and nonhuman	-.027 (.242)	- inf – .101	-.538 (23)	<i>n.s.</i>	.112
		Faces and Bodies	.027 (.172)	- inf – .118	.773 (23)	<i>n.s.</i>	.157
		Natural and Artificial	-.0003 (.153)	- inf – .081	.010 (23)	<i>n.s.</i>	.002
		Big and Small	-.002 (.163)	- inf – .088	.046 (23)	<i>n.s.</i>	.012
1b	6 m	Eight-categories	-.020 (.113)	- inf – .040	-.847 (23)	.203	.177
		Animacy	-.041 (.095)	- inf – .009	-2.139 (23)	.022	.432
		Human and nonhuman	-.043 (.243)	- inf – .085	-.865 (23)	.198	.177
		Faces and Bodies	-.055 (.195)	- inf – .048	-1.375 (23)	.091	.282
		Natural and Artificial	.051 (.177)	- inf – .144	1.405 (23)	<i>n.s.</i>	.288
		Big and Small	-.009 (.250)	- inf – .123	-.170 (23)	<i>n.s.</i>	.036
2	8 m	Eight-categories	-.119 (.083)	- inf – -.076	-7.033 (23)	<.0001	1.434
		Animacy	-.111 (.098)	- inf – -.060	-5.556 (23)	<.0001	1.133
		Human and nonhuman	-.057 (.142)	- inf – .018	-1.976 (23)	.030	.401
		Faces and Bodies	-.070 (.216)	- inf – .044	-1.581 (23)	.064	.324
		Natural and Artificial	.007 (.142)	- inf – .082	.229 (23)	<i>n.s.</i>	.049
		Big and Small	-.093 (.162)	- inf – -.007	-2.806 (23)	.005	.574
3	8 m	Eight-categories	-.049 (.458)	- inf – .008	-2.233 (23)	.018	.458
		Animacy	-.051 (.111)	- inf – .008	-2.243 (23)	.017	.460
		Human and nonhuman	-.076 (.180)	- inf – .019	-2.078 (23)	.025	.422
		Faces and Bodies	-.089 (.233)	- inf – .034	-1.869 (23)	.037	.382
		Natural and Artificial	.090 (.152)	- inf – .170	2.880 (23)	<i>n.s.</i>	.592
		Big and Small	.056 (.198)	- inf – .160	1.389 (23)	<i>n.s.</i>	.283

Note: Exp., experiment; m, months; CA, corrected age; CI, 99.17% confidence interval; Highlighted in bold are the significant results; *a* = .0083, one-tailed; *n.s.* = non-significant results with *P* > .250.

3. Analysis of mean looking times (Exps. 1-3)

Separate one-way ANOVAs for 6-month-olds in Exp. 1a-b, for 8-month-olds in Exp. 2 and for 8-month-olds in Exp. 3 revealed the effect of categories on mean looking times (MLTs). Following *t*-tests were carried out to follow up on the effect of category.

For 6-month-old infants in Exp. 1a, pairwise comparisons (one sample *t*-tests, $\alpha_{\text{corrected}} = 0.0018$, two-tailed) revealed that infants looked longer at human faces relative to all other categories (*P*s < 0.0001 for all of the 7 comparisons between the human face category and other categories respectively). Within the inanimate categories, infants looked longer at big artificial objects than at small ones (*P* <

0.001), and in general at big than small objects, although comparisons do not survive the multiple comparisons correction ($P_s < .03$ except for the comparison between big and small natural objects for which $P < .27$). For 6-month-old infants in Exp. 1b, results revealed that they prefer (i.e., looked longer at) human faces *vs.* all other categories (all $P_s < 0.0001$) as well as animate *vs.* inanimate (10 out of 16 comparisons are significant with $P_s < .001$; 3 additional comparisons would be significant without the multiple tests correction), big *vs.* small (natural or artificial) inanimate objects (3 out of 4 comparisons are significant with $P_s < .0001$; the remaining does not survive the multiple tests correction but would be significant without it, $P = .004$), for humans compare to nonhuman animals (3 out of 4 comparisons are significant with $P_s < .0001$; the remaining one would be without the multiple tests correction, $P = .014$) as well as for faces compare to bodies (all $P_s < .0001$). Unlike in Exp. 1a, the preference for human faces and big objects in Exp. 1b was not conflated with the preference for large images, as all images featured the same number of pixels. However, the mean looking time for individual images, averaged across participants, negatively correlated with the image shape elongation ($\rho = -0.407$, $P < 0.001$), and positively correlated with the image compactness ($\rho = 0.603$, $P < 0.0001$): the less elongated and the more compact the shape, the longer the looking time (see also Supplementary Results 4). See Supplementary Table 4 for t -values and P -values of all the pairwise comparisons reported here.

For 8-month-old infants in Exp. 2, pairwise comparisons (one sample t -tests, $\alpha_{\text{corrected}}$: 0.0018) revealed longer looking times towards animate than inanimate objects (11 out of 16 comparisons were significant; 3 of the 5 remaining comparisons showed non-significant trends, that is, effects that did not reach the significance level, corrected for multiple comparisons). A t -test comparing all animate *vs.* all inanimate categories confirmed this effect (animate: $M \pm SD = 2.066 \text{ s} \pm 0.317$; inanimate: $M \pm SD = 1.483 \text{ s} \pm 0.234$; $M_{\text{difference}} \pm SD = 0.583 \pm 0.352$; 95% CI = 0.434 – 0.731; $t(23) = 8.120$, $P < 0.0001$; $d = 1.656$). Comparisons between faces and bodies categories showed mostly significant difference (3 out of 4 comparisons were significant), which can be summarized as a preference for faces categories ($M \pm SD = 2.296 \text{ s} \pm 0.428$) over bodies ($M \pm SD = 1.836 \text{ s} \pm 0.328$; $M_{\text{difference}} \pm SD = 0.461 \pm 0.423$; 95% CI = 0.282 – 0.639; $t(23) = 5.331$, $P < 0.0001$; $d = 1.090$). Among the inanimate categories, only the comparison between artificial-big *vs.* artificial-small showed a non-significant trend ($P = 0.005$; for other comparisons all $P_s > 0.10$). The MLTs analysis also revealed a preference for human faces over almost all other categories, except over nonhuman animal faces (all $P_s < .001$ except for the comparison with nonhuman animal faces where $P = .085$).

Eight-month-old preterm infants in Exp. 3 looked longer at human faces *vs.* all other categories as well (all $P_s < 0.0001$). Comparisons between faces and bodies categories showed mostly significant

difference (3 out of 4 comparisons were significant), which can be summarized as a preference for faces categories ($M \pm SD = 2.098 \text{ s} \pm 0.360$) over bodies ($M \pm SD = 1.564 \text{ s} \pm 0.362$; $M_{\text{difference}} \pm SD = 0.534 \pm 0.353$; 95% CI = 0.385 – 0.683; $t(23) = 7.417$, $P < 0.0001$; $d = 1.513$).

Supplementary Table 4. Absolute t-values and P-values of post-hoc t-test between MLTs of the eight categories

Exp.	Age	Categories																								
		HB	NHF	NHB	NB	NS	AB	AS	P																	
1	8 m	HF	t	4.145	P	<0.0001	t	1.803	P	<0.001	t	9.085	P	<0.0001	t	6.544	P	<0.0001	t	5.620	P	<0.0001	t	7.620	P	<0.0001
		HB	t	4.145	P	<0.0001	t	1.803	P	<0.001	t	9.085	P	<0.0001	t	6.544	P	<0.0001	t	5.620	P	<0.0001	t	7.620	P	<0.0001
		NHF	t	1.654	P	0.112	t	1.094	P	n.s.	t	3.863	P	<0.001	t	3.754	P	0.013	t	2.695	P	0.013	t	5.411	P	<0.0001
		NHB	t	4.317	P	<0.001	t	4.317	P	<0.001	t	6.094	P	<0.0001	t	5.550	P	<0.0001	t	3.998	P	<0.001	t	5.771	P	<0.0001
1	6 m	HF	t	7.115	P	<0.0001	t	6.129	P	<0.0001	t	6.660	P	<0.0001	t	7.167	P	<0.0001	t	5.896	P	<0.0001	t	6.236	P	<0.0001
		HB	t	7.115	P	<0.0001	t	6.129	P	<0.0001	t	6.660	P	<0.0001	t	7.167	P	<0.0001	t	5.896	P	<0.0001	t	6.236	P	<0.0001
		NHF	t	0.316	P	n.s.	t	0.154	P	n.s.	t	0.251	P	n.s.	t	0.860	P	n.s.	t	1.741	P	0.095	t	2.228	P	0.036
		NHB	t	0.377	P	n.s.	t	0.377	P	n.s.	t	0.000	P	n.s.	t	1.003	P	n.s.	t	1.241	P	0.227	t	1.926	P	0.067
2	6 m	HF	t	12.328	P	<0.0001	t	7.264	P	<0.0001	t	11.330	P	<0.0001	t	10.339	P	<0.0001	t	9.349	P	<0.0001	t	12.792	P	<0.0001
		HB	t	12.328	P	<0.0001	t	7.264	P	<0.0001	t	11.330	P	<0.0001	t	10.339	P	<0.0001	t	9.349	P	<0.0001	t	12.792	P	<0.0001
		NHF	t	7.439	P	<0.0001	t	2.665	P	0.014	t	6.939	P	<0.0001	t	3.285	P	0.003	t	9.641	P	<0.0001	t	1.489	P	0.150
		NHB	t	4.313	P	<0.001	t	6.968	P	<0.0001	t	4.313	P	<0.001	t	5.582	P	<0.0001	t	2.081	P	0.049	t	8.152	P	<0.0001
3	8 m	HF	t	6.386	P	<0.0001	t	6.010	P	<0.0001	t	7.959	P	<0.0001	t	9.066	P	<0.0001	t	5.808	P	<0.0001	t	10.335	P	<0.0001
		HB	t	6.386	P	<0.0001	t	6.010	P	<0.0001	t	7.959	P	<0.0001	t	9.066	P	<0.0001	t	5.808	P	<0.0001	t	10.335	P	<0.0001
		NHF	t	0.398	P	n.s.	t	1.825	P	0.082	t	0.590	P	n.s.	t	1.790	P	0.088	t	0.252	P	n.s.	t	3.108	P	0.005
		NHB	t	3.872	P	<0.001	t	3.872	P	<0.001	t	1.050	P	n.s.	t	2.613	P	0.016	t	0.860	P	n.s.	t	4.262	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t	4.250	P	0.001	t	2.130	P	0.055	t	3.665	P	0.003	t	5.108	P	<0.001
3	10 m	HF	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		HB	t	1.370	P	0.196	t	0.116	P	n.s.	t	0.741	P	n.s.	t	1.174	P	n.s.	t	2.772	P	0.017	t	2.779	P	0.017
		NHF	t	1.582	P	0.140	t	1.582	P	0.140	t	0.364	P	n.s.	t	0.205	P	n.s.	t	1.933	P	0.077	t	2.898	P	0.013
		NHB	t	2.461	P	0.030	t	0.848	P	n.s.	t</															

4. Relationship between MLTs and low-level features of images.

Because 6- and 8-month-olds' looking behavior, whether full-term or preterm (Exps. 1-3), appeared to be guided by low-level features such as the size of images, the compactness or the elongation (Supplementary Fig. 1), we tested whether the categories of images that infants looked at for a longer time, corresponded to the categories including larger images, more compact and/or less elongated objects. To this end, we computed the MLTs, the mean number of pixels and the mean elongation ratio for each category (Supplementary Fig. 2). The three preferred categories, human faces and natural/artificial big objects, constituted the largest, the least elongated and the more compact images. Furthermore, mean looking times (MLTs) correlated significantly with size, compactness and elongation in Exp. 1a for 6-month-olds, with compactness and elongation in Exp. 1b for 6-month-olds (see Main text), with compactness in Exp. 2 for 8-month-olds and with size and compactness for 8-month-old preterm infants in Exp. 3.

A Exp. 1a – Subject 6 – 6 months



B Exp. 1b – Subject 2 – 6-months



C Exp. 2 – Subject 8 – 8 months



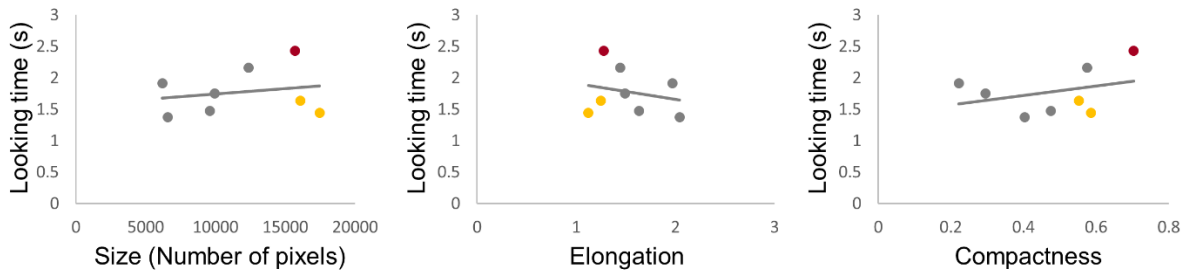
D Exp. 2 – Subject 5 – Preterm 8 months



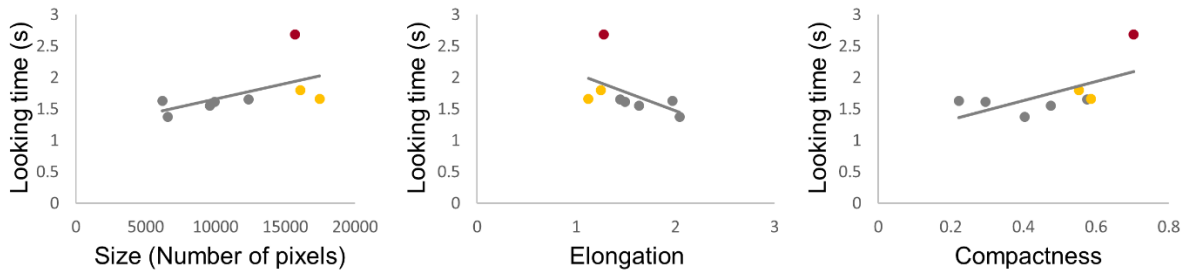
Signed matrix Size's matrix Elongation's matrix Compactness' matrix Color's matrix
 Very dissimilar — Very similar Very dissimilar Very similar Very dissimilar

Supplementary Fig. 1. Signed DLT-RDM (left), size RDM (middle left), elongation RDM (middle), compactness (middle right) and color RDM (right) for 6-month-olds in Experiment 1a and b, 8-month-olds in Experiment 2 and 8-month-old preterm infants in Experiment 3. Example of a signed DLT-RDM, a size RDM, an elongation RDM, a compactness RDM and a color RDM for one subject of the 6-month-old group in Experiment 1a (a), for one subject of the 6-month-old group in Experiment 1b (b), for one subject of the 8-month-old group in Experiment 2 (c) and for one subject of the 8-month-old preterm infants group in Experiment 3 (d).

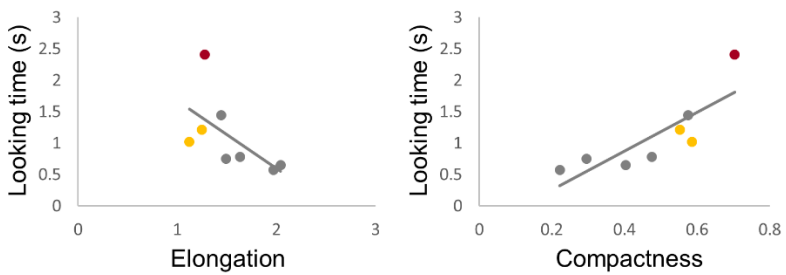
A 8-month-olds – Exp. 1



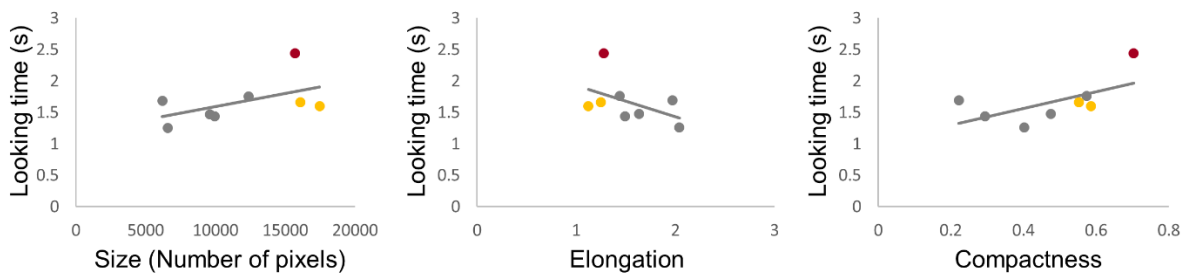
B 6-month-olds – Exp. 1



C 6-month-olds – Exp. 2



D 8-month-old preterm infants – Exp. 3



● Human face category ● Natural and Artificial big categories

Supplementary Fig. 2. Looking times as a function of image size (left), elongation (middle) and compactness (right) for 6-month-olds in Experiment 1a and b, for 8-month-olds in Experiment 2 and for 8-month-old preterm infants in Experiment 3. Relationship between looking times (s) and average size of the images (number of pixels) for each category, average elongation (ratio between width and height) for each category and average compactness for 6-month-olds in experiment 1a (a), relationship between looking times (s) for each category and average elongation and compactness for 6-month-olds in Experiment 1b (b), relationship between looking times (s) for each category and average size of the images, average elongation and average compactness for 8-month-olds in Experiment 2 (c), and for 8-month-old preterm infants in Experiment 3 (d).

Given these results, we performed a novel stepwise linear regression analysis and DLTs analysis after removing the variance explained by compactness, size and elongation for 6-month-old (Exp. 1a) and for 8-month-old (Exp. 2) full-term infants as well as for 8-month-old preterm infants (Exp. 3; Supplementary Tables 5-6), and the variance explained by elongation and compactness for the second group of 6-month-old in Exp. 1b (Supplementary Tables 4- 5).

Supplementary Table 5. Results of the stepwise linear regression with artificial models, removing variance explained by size, compactness and elongation for 6- and 8-month-old infants' DLT-RDM in Experiments 1-3, and by compactness and elongation for 6-month-old infants' DLT-RDM in Experiment 1

Exp.	Age	Regressor	Mean β (<i>SD</i>)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1a	6 m	Eight-categories	-.037 (.116)	-.105 – .032	-1.549 (23)	.135	.319
		Animacy	.036 (.114)	-.031 – .103	1.551 (23)	.135	.316
		Humanness	.042 (.263)	-.113 – .196	.774 (23)	<i>n.s.</i>	.160
		Faces/Bodies	-.094 (.198)	-.211 – .022	-2.338 (23)	.028	.475
		Natural/Artificial	.005 (.221)	-.126 – .135	.102 (23)	<i>n.s.</i>	.023
		Big/Small	-.050 (.146)	-.136 – .036	-1.682 (23)	.106	.342
1b	6 m	Eight-categories	-.043 (.114)	-.110 – .024	-1.846 (23)	.078	.377
		Animacy	.019 (.101)	-.041 – .078	.901 (23)	<i>n.s.</i>	.188
		Humanness	.018 (.218)	-.111 – .146	.392 (23)	<i>n.s.</i>	.083
		Faces/Bodies	-.075 (.210)	-.199 – .049	-1.755 (23)	.093	.357
		Natural/Artificial	-.002 (.214)	-.127 – .124	-.033 (23)	<i>n.s.</i>	.009
		Big/Small	-.018 (.264)	-.174 – .137	-.343 (23)	<i>n.s.</i>	.068
2	8 m	Eight-categories	.070 (.080)	.023 – .117	4.301 (23)	<.001	.875
		Animacy	.111 (.114)	.044 – .179	4.780 (23)	<.0001	.974
		Humanness	.071 (.174)	-.031 – .173	2.001 (23)	.057	.408
		Faces/Bodies	.045 (.181)	-.062 – .151	1.208 (23)	.239	.249
		Natural/Artificial	.044 (.177)	-.060 – .148	1.220 (23)	.235	.249
		Big/Small	.064 (.180)	-.042 – .170	1.736 (23)	.096	.356
3	8 m	Eight-categories	-.049 (.105)	-.013 – .111	2.292 (23)	.031	.467
		Animacy	.060 (.082)	.012 – .108	3.571 (23)	.002	.732
		Humanness	.077 (.237)	-.063 – .217	1.597 (23)	<i>n.s.</i>	.325
		Faces/Bodies	.037 (.197)	-.080 – .153	.907 (23)	.038	.188
		Natural/Artificial	-.103 (.209)	-.226 – .021	-2.398 (23)	.177	.493
		Big/Small	-.015 (.205)	-.135 – .106	-.349 (23)	<i>n.s.</i>	.073

Note: Exp., experiment; m, months; CA, corrected age; CI, confidence interval; Highlighted in bold are the significant results; *a* = .0083, two-tailed; *n.s.* = non-significant results with *P* > .250.

Supplementary Table 6. Results of the DLTs analysis of within versus between category comparisons, removing variance explained by size, compactness and elongation for 6- and 8-month-old infants in Experiments 1 and 3 and by compactness and elongation for 6-month-old infants in Experiment 2

Exp.	Age	Comparisons within/between	Mean of the difference (SD)	CI (min – max)	<i>t</i> (df)	<i>P</i>	Cohen's <i>d</i>
1a	6 m	Diagonal	.032 (.104)	- inf – .087	1.487 (23)	<i>n.s.</i>	.308
		Animate and inanimate	-.030 (.105)	- inf – .026	-1.375 (23)	.091	.286
		Human and nonhuman	-.032 (.280)	- inf – .116	-.558 (23)	<i>n.s.</i>	.114
		Faces and Bodies	.108 (.199)	- inf – .213	2.644 (23)	<i>n.s.</i>	.543
		Natural and Artificial	.010 (.212)	- inf – .122	.231 (23)	<i>n.s.</i>	.047
		Big and Small	.063 (.146)	- inf – .141	2.117 (23)	<i>n.s.</i>	.432
1b	6 m	Diagonal	.042 (.119)	- inf – .104	1.729 (23)	<i>n.s.</i>	.353
		Animate and inanimate	-.019 (.104)	- inf – .036	-0.872 (23)	.196	.183
		Human and nonhuman	.0005 (.203)	- inf – .107	-0.012 (23)	<i>n.s.</i>	.002
		Faces and Bodies	.086 (.217)	- inf – .201	1.951 (23)	<i>n.s.</i>	.396
		Natural and Artificial	.018 (.219)	- inf – .134	.412 (23)	<i>n.s.</i>	.082
		Big and Small	.023 (.267)	- inf – .164	.418 (23)	<i>n.s.</i>	.086
2	8 m	Diagonal	-.108 (.101)	- inf – -.054	-5.221 (23)	<.0001	1.069
		Animate and inanimate	-.112 (.113)	- inf – -.053	-4.863 (23)	<.0001	.991
		Human and nonhuman	-.074 (.196)	- inf – .029	-1.847 (23)	.039	.378
		Faces and Bodies	-.048 (.190)	- inf – .053	-1.229 (23)	.116	.253
		Natural and Artificial	-.029 (.185)	- inf – .069	-.761 (23)	.227	.157
		Big and Small	-.053 (.175)	- inf – .040	-1.468 (23)	.078	.303
3	8 m	Diagonal	.041 (.188)	- inf – .021	-1.701 (23)	.051	.348
		Animate and inanimate	-.062 (.085)	- inf – -.017	-3.581 (23)	<.001	.729
		Human and nonhuman	-.087 (.266)	- inf – .053	-1.608 (23)	.061	.327
		Faces and Bodies	-.047 (.212)	- inf – .065	-1.082 (23)	.145	.222
		Natural and Artificial	.108 (.189)	- inf – .208	2.803 (23)	<i>n.s.</i>	.571
		Big and Small	.034 (.219)	- inf – .150	.753 (23)	<i>n.s.</i>	.155

Note: Exp., experiment; m, months; CA, corrected age; CI, confidence interval; Highlighted in bold are the significant results; $\alpha = .0083$, two-tailed; *n.s.* = non-significant results with $P > .250$

5. Experiment 1: Data analysis with more lenient inclusion criteria

The attrition rate for the group of 6-month-olds in Exp. 1b was larger compared to the group of 6-month-olds in Exp. 1a (8% versus 64% in Exp. 2). Data from 42 out of 66 tested infants were discarded in Exp. 1b, while data from 2 out of 26 tested infants were discarded in Exp. 1a. A larger attrition rate may lead to the inclusion of more attentive infants, which could explain the different results in Exp. 1b compared to Exp. 1a.

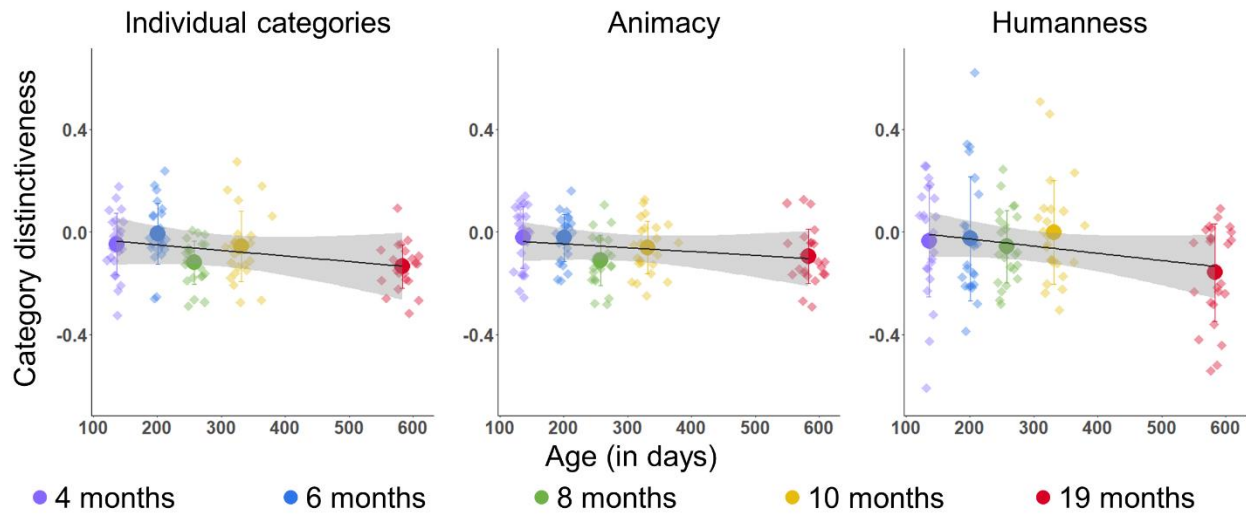
To take into account this possibility, we ran a novel analysis on the data from Exp. 1b, using inclusion criteria that yielded an attrition rate comparable to that of 6-month-olds in Exp. 1a. In particular, we modified the criterion for trial inclusion, selecting all trials in which infants looked at

images for at least 300 ms (instead of 1000 ms in the original analysis reported in the main text). This change yielded the exclusion of 6 out of 66 infants (attrition rate of 9%). The analysis of this dataset confirmed the results reported in the main text. The stepwise linear regression analyzing the structure of the absolute RDMs did not identify significant regressor ($\alpha_{\text{corrected}} = 0.0083$), although when looking at the correlation between the animacy model and infants' matrices, we found a significant correlation ($M_{\text{Fisher}} \pm \text{SD} = 0.056 \pm 0.151$; 95% CI = 0.017 – 0.095; $t(59) = 2.874$; $P = 0.006$; $d = 0.371$). The DLTs analysis however showed a significant difference between within-category and between-category DLTs for the animacy model ($M_{\text{difference}} = -0.036$; 99.17% CI = -inf – -0.004; $t(23) = -2.778$, $P = 0.004$; $d = 0.364$) but no other models ($ts < 2.190$; $Ps > 0.01$).

6. Analysis of category distinctiveness from 4 to 19 months of age

In addition to the previous analysis, we also explored the relation between infants' age and their category distinctiveness (Nordt et al., 2023), including full-term infants tested in chapters 3 and 4. The category distinctiveness is defined as the difference between the within-category dissimilarity and the between-category dissimilarity. Thus, a negative score of distinctiveness reveals a smaller within-category dissimilarity compare to the between-category dissimilarity, so a better distinctiveness of the category. We assessed the existence of a linear relationship between infants' age and infants' category distinctiveness for the 8-individual categories, the animacy (animate-inanimate) and the humanness (humans-nonhuman animals) categories. For the 8-individual categories distinctiveness, we averaged, for each infant, the dissimilarity of the 8 diagonal values of the infant's DLT-RDM as the within-category dissimilarity, and all other values off the diagonal as the between-category dissimilarity. For the animacy distinctiveness, we averaged the dissimilarity within all animate-animate and inanimate-inanimate comparisons of the infant's DLT-RDM as the within-category dissimilarity, and the dissimilarity between animate-inanimate comparisons as the between-category dissimilarity. For the humanness distinctiveness, we averaged the dissimilarity within all human-human and nonhuman animal-nonhuman animal comparisons of the infant's DLT-RDM as the within-category dissimilarity, and the dissimilarity between human-nonhuman animal comparisons as the between-category dissimilarity. We then subtracted the between-category dissimilarity to the within-category dissimilarity, and evaluated a linear relationship between infants' category distinctiveness and infants' age.

Results show a linear relationship between age and the distinctiveness of the 8-individual categories ($F(1,119) = 10.369, P = .002$; Supplementary Fig. 3), animacy ($F(1,119) = 5.516, P = .020$)



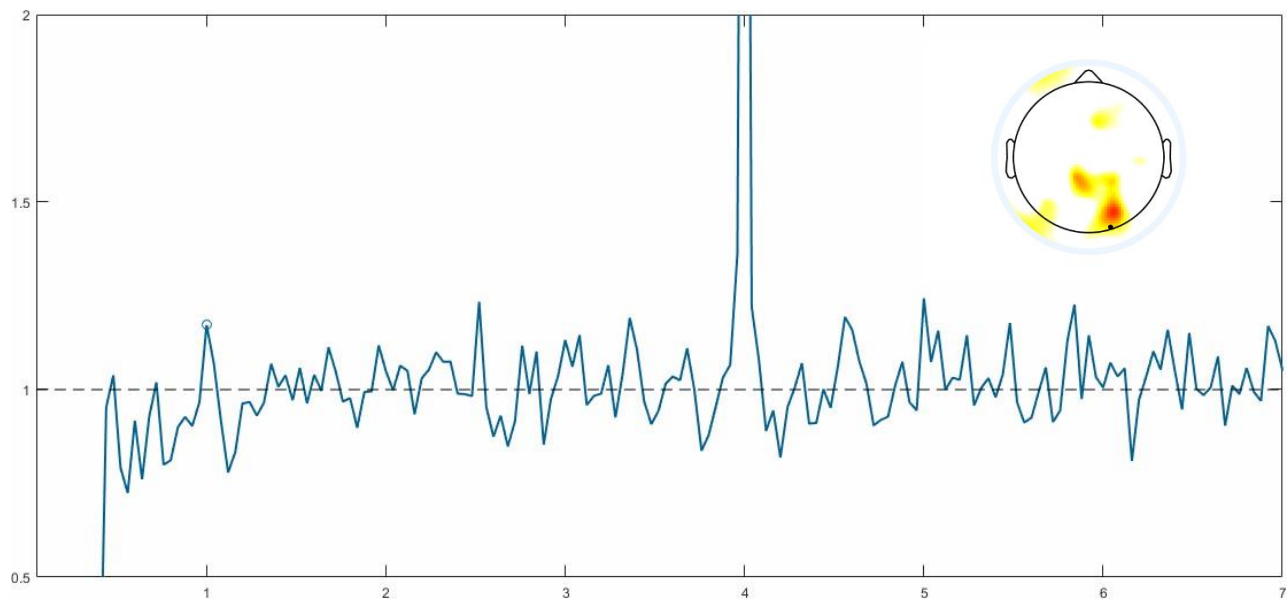
Supplementary Fig. 3. Evolution of infants' category distinctiveness through age. Category distinctiveness increases as infants grow older, for the 8-individual categories, the Animacy and the Humanness categories. Age are express in days, and group age are represented by colors.

and humanness ($F(1,119) = 5.786, P = .018$). Altogether, this analysis suggests an increase in distinctiveness of categories with age, from 4 to 19 months of age.

9.3. Chapter 6 – Supplementary results

In addition to study the role of low-level visual features in adults, we also studied whether infants' categorization of animate and inanimate objects, as captured in Chapter 5, could also be captured with only lower-level visual features, just like in adults. We adapted the experiment to 4-month-old infants, the youngest group for which we found categorization between animate and inanimate, presenting the images at 4 Hz, and the categorical stimulation as 1 Hz (1 tagged image every 4 images). We used the color phase-scrambled stimuli set, to make sure infants could easily see them on the screen, and would be willing to look at it.

We found a significant brain response over one occipital electrode for 4-month-old infants looking at phase-scramble images ($M_{Amplitude} \pm sd = 0.223 \pm 0.598$; 95% CI = 0.044 – Inf; $t(31) = 2.112$; $P = .021$; $d = 0.373$).



Supplementary Figure 1. SNR and topoplot. Infants Signal-to-noise ratio (SNR) and distribution over the scalp (topoplot) of the response for the animate-inanimate categorization of phase-scrambled images.