



**HAL**  
open science

# Attractivité des visages chez les mandrills et les humains : apport de l'intelligence artificielle générative et prédictive

Nicolas Dibot

► **To cite this version:**

Nicolas Dibot. Attractivité des visages chez les mandrills et les humains : apport de l'intelligence artificielle générative et prédictive. Ecologie, Environnement. Université de Montpellier, 2024. Français. NNT : 2024UMONG018 . tel-04895020

**HAL Id: tel-04895020**

**<https://theses.hal.science/tel-04895020v1>**

Submitted on 17 Jan 2025

**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.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Ecologie, Evolution, Ressources Génétiques, Paléobiologie (EERGP)

École doctorale GAIA – Biodiversité Agriculture Alimentation Environnement Terre Eau

Unités de recherche :

Centre d'Ecologie Fonctionnelle et Evolutive (CEFE)

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM)

## Attractivité des visages chez les mandrills et les humains : Apport de l'intelligence artificielle générative et prédictive

Présentée par Nicolas DIBOT

Le 24 septembre 2024

Sous la direction de Julien RENOULT et William PUECH

Devant le jury composé de

Jenny BENOIS-PINEAU, Professeure des universités, Laboratoire Bordelais de Recherche en Informatique	Rapportrice, Présidente du jury
Emmanuelle POUYDEBAT, Directrice de recherche, Muséum National d'Histoire Naturelle	Rapportrice
Alexandra ALVERGNE, Chargée de recherche, Institut des Sciences de l'Evolution de Montpellier	Examinatrice
Nicolas CLAUDIERE, Chargé de recherche, Centre de Recherche en Psychologie et Neurosciences	Examineur
Julien RENOULT, Chargé de recherche, Centre d'Ecologie Fonctionnelle et Evolutive	Directeur de thèse
William PUECH, Professeur des universités, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier	Directeur de thèse
Marie CHARPENTIER, Directrice de recherche, Institut des Sciences de l'Evolution de Montpellier	Encadrante (invitée)



UNIVERSITÉ DE  
MONTPELLIER



*Une thèse, c'est des problèmes.*





# Remerciements

Dans le cadre de ces travaux de thèse, je souhaite remercier les personnes qui y ont contribué.

Je remercie en tout premier lieu l'équipe de direction de thèse pour l'encadrement tout au long de ces trois ans : Julien Renoult, William Puech et Marie Charpentier.

Je remercie aussi les chercheuses et chercheurs qui ont accepté de faire partie de mon jury : Emmanuelle Pouydebat, Jenny Benois-Pineau, Alexandra Alvergne et Nicolas Claidière.

Je souhaite aussi remercier les membres du comité de thèse : Michel Raymond, Anne Charmantier, Gérard Subsol et Philippe Montesinos.

J'adresse mes remerciements aux personnes avec qui j'ai pu collaborer ou échanger de près ou de loin dans le cadre de ces travaux, et tout particulièrement Alice Baniel, Yseult Héjja-Brichard, Thierry Tsoumbou, Brice Ndinga, Jean Nzue Nguema, Melvin Bardin et Roland Bertin-Johannet.

Je tiens à remercier les collègues du labo avec qui j'ai pu passer de bons moments. Il serait difficile de citer tout le monde sans oublier personne. Néanmoins, une mention spéciale pour Wakinyan.

Enfin, je remercie toutes les personnes de mon entourage familial et amical, en particulier mes parents.

Merci à tous !



# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Table des matières</b>	<b>v</b>
<b>1 Avant-propos</b>	<b>1</b>
<b>INTRODUCTION GÉNÉRALE</b>	<b>3</b>
<b>2 Biologie et psychologie de l'attractivité faciale</b>	<b>5</b>
2.1 Qu'est ce que l'attractivité faciale? . . . . .	5
2.1.1 Perception visuelle . . . . .	5
2.1.2 Mécanismes de l'attractivité des visages . . . . .	6
2.2 Perception et beauté . . . . .	8
2.2.1 Généralités . . . . .	8
2.2.2 Fluence et prototypicalité . . . . .	9
2.3 Sélection sexuelle et choix de partenaire . . . . .	10
2.3.1 Définitions et généralités . . . . .	10
2.3.2 Dimorphisme sexuel et impacts . . . . .	11
<b>3 L'intelligence artificielle pour modéliser l'attractivité faciale</b>	<b>13</b>
3.1 Intelligence artificielle . . . . .	14
3.1.1 Apprentissage automatique . . . . .	14
3.1.2 Réseaux de neurones artificiels et apprentissage profond . . . . .	16
3.1.3 Intelligence artificielle générative . . . . .	20
3.2 Un modèle et des métriques . . . . .	22
<b>4 L'intelligence artificielle pour comprendre expérimentalement l'attractivité faciale</b>	<b>27</b>
4.1 L'approche expérimentale . . . . .	27
4.1.1 Généralités . . . . .	27
4.1.2 Ecologie sensorielle et écologie visuelle . . . . .	28
4.1.3 Construire un protocole pour tester une préférence visuelle . . . . .	28
4.2 L'intelligence artificielle générative comme outil expérimental . . . . .	30
4.2.1 Fabriquer des stimuli . . . . .	31
4.2.2 Apport de l'intelligence artificielle générative . . . . .	32
<b>5 Des modèles pour étudier l'attractivité faciale à l'aune de l'intelligence artificielle</b>	<b>35</b>
5.1 Force et complémentarités des modèles humains et mandrills . . . . .	35
5.2 Les mandrills : choix de partenaire, signaux visuels et populations étudiées . . . . .	37
5.2.1 Ecologie et organisation sociale des mandrills . . . . .	37
5.2.2 Signaux visuels, perception des visages et choix de partenaire chez les mandrills . . . . .	37
5.2.3 Le Projet Mandrillus, la base de données de visages de mandrills et les travaux de vision par ordinateur chez les mandrills . . . . .	37
5.2.4 Des mandrills en semi-captivité dans un centre de recherche au Gabon . . . . .	40
<b>6 Objectifs de la thèse</b>	<b>41</b>



<b>CHAPITRE 1 : GENERATION AND EDITING OF MANDRILL FACES : APPLICATION TO SEX EDITING AND ASSESSMENT</b>	<b>43</b>
<b>CHAPITRE 2 : ATTRACTIVITY OF FACIAL FEMINITY OF A NON-HUMAN PRIMATE : EXPERIMENTAL EVIDENCE BASED ON GENERATIVE AI</b>	<b>69</b>
<b>CHAPITRE 3 : SPARSITY IN AN ARTIFICIAL NEURAL NETWORK PREDICTS BEAUTY : TOWARDS A MODEL OF PROCESSING-BASED AESTHETICS</b>	<b>87</b>
<b>DISCUSSION GÉNÉRALE</b>	<b>107</b>
<b>7 Préambule à la discussion générale</b>	<b>109</b>
<b>8 L'origine des préférences</b>	<b>111</b>
<b>9 Évaluer expérimentalement une préférence en écologie comportementale</b>	<b>113</b>
9.1 Regard critique sur le dispositif expérimental . . . . .	113
9.2 Les limites du paradigme du temps d'observation . . . . .	114
9.3 Analyse complémentaires et poursuite de ces travaux . . . . .	115
<b>10 L'IA en écologie et en évolution</b>	<b>117</b>
10.1 Encoder des données réelle dans un espace latent . . . . .	117
10.2 L'IA générative en écologie et évolution . . . . .	119
10.2.1 Augmenter et compléter des données . . . . .	119
10.2.2 Créer des stimuli . . . . .	120
10.2.3 Simuler l'évolution . . . . .	120
10.2.4 Prédire le vivant . . . . .	121
<b>11 Conclusion</b>	<b>123</b>
<b>APPENDIX</b>	<b>125</b>
<b>Bibliographie</b>	<b>153</b>
<b>Résumé</b>	<b>165</b>

Les travaux de thèse sur l'attractivité des visages que nous allons présenter ici se placent autour de la question de l'origine des préférences. En particulier, nous avons exploré la dualité entre la préférence en tant qu'indicateur de qualité ou en tant que biais perceptuel. Pour ce faire, nous avons étudié des cas d'applications concrets permettant d'éclairer l'importance respective de ces deux explications, constituant les 3 objectifs de cette thèse (Cf. Fig. 1.1).

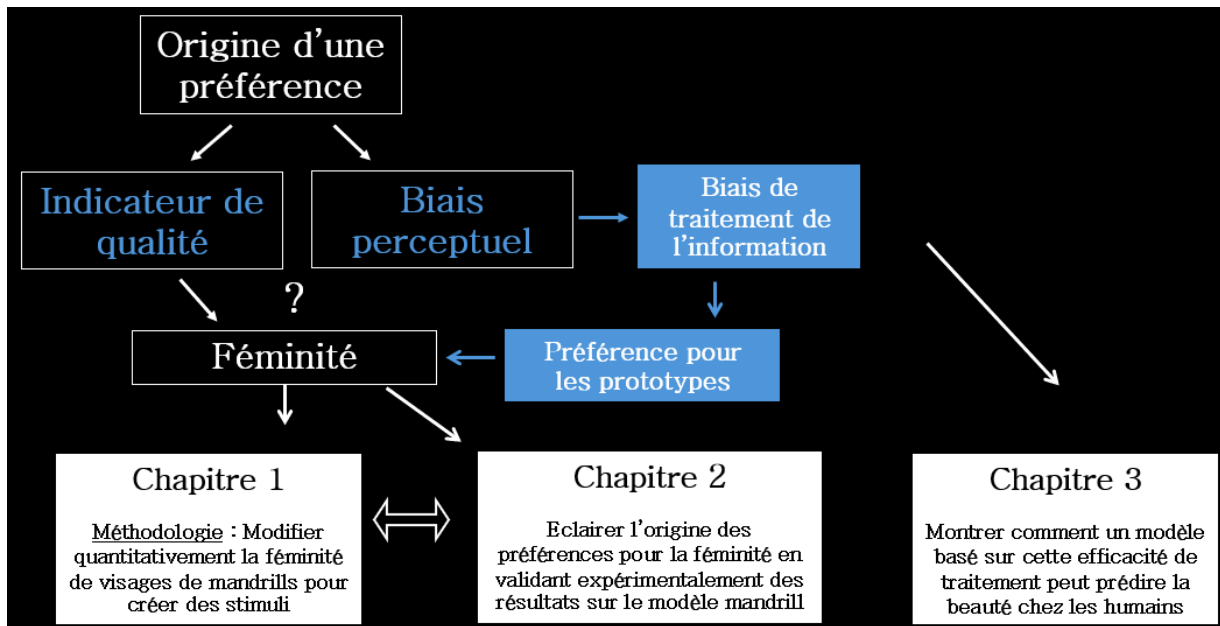


FIG. 1.1 : Articulations des 3 objectifs de la thèse autour de la question de l'origine des préférences (*production personnelle*)

Avant de détailler les enjeux et réalisations issus de ces trois objectifs, nous explorerons en introduction générale les différents concepts nécessaires pour comprendre l'articulation de nos travaux autour de l'attractivité faciale et les choix des approches variées utilisées. Ensuite, nous présenterons ces travaux sous la forme d'articles scientifiques. Enfin, nous reviendrons sur les perspectives et les limites de ces travaux dans une discussion générale, où nous développerons aussi des axes de travail qui n'auront pas été traités dans les chapitres précédents.



# **INTRODUCTION GÉNÉRALE**



# Biologie et psychologie de l'attractivité faciale

# 2

## 2.1 Qu'est ce que l'attractivité faciale ?

Les visages ont de nombreuses fonctions essentielles dans les échanges sociaux, que ce soit chez les humains ou les autres animaux. Ces fonctions peuvent concerner le choix de partenaire sexuel ou amoureux, ou la prise de décision dans un cadre collaboratif, conflictuel, ou encore professionnel. Pour des espèces sociales, l'attractivité du visage est un paramètre rapidement pris en compte par les congénères et qui va donc façonner les interactions avec cet individu.

Afin de comprendre comment un visage est perçu, nous développerons en premier lieu le fonctionnement et l'organisation du système visuel. Ensuite, nous décrirons quels mécanismes entrent en jeu dans l'évaluation de l'attractivité des visages.

### 2.1.1 Perception visuelle

Un tiers des embranchements animaux ont des yeux [1], et d'autres règnes possèdent aussi des organes sensibles à la lumière. Le système visuel est l'ensemble des organes d'un être vivant doté de vision permettant l'acquisition, le traitement et l'interprétation d'information provenant de rayons lumineux.

Chez les mammifères, ces rayons lumineux pénètrent tout d'abord dans l'œil par la cornée, transparente, située à l'avant, puis par le cristallin, une lentille biconvexe permettant de focaliser la lumière. La pupille, contrôlée par l'iris, permet ensuite de plus ou moins laisser entrer la lumière sur la rétine, couche sensible à la lumière. L'information visuelle, jusqu'à maintenant sous forme de rayons lumineux, change alors de nature. Les photorécepteurs situés sur la rétine, cône et bâtonnets, responsables respectivement de la vision des couleurs et de la vision à faible luminosité, transforment l'information visuelle en signaux électriques.

Les fibres nerveuses du nerf optique transmettent ces signaux au cerveau, en particulier au cortex visuel, situé dans le lobe occipital du cerveau (à l'arrière du crâne). C'est la zone du cerveau responsable du traitement des informations visuelles. Le cortex visuel est composé de plusieurs aires. L'aire V1, aussi connue sous le nom de cortex visuel primaire, traite les informations visuelles simples et localisées, comme les contrastes de luminosité ou les contours et les bords des objets. Les aires V2 à V5 forment le cortex visuel associatif. Celui-ci est spécialisé dans l'intégration et l'interprétation d'informations visuelles plus complexes comme les mouvements, les couleurs ou les textures. Ces aires sont organisées hiérarchiquement. Les signaux traités dans l'aire V1 sont ensuite transmis aux aires V2 et V3, où sont traités les contours et les textures, puis aux aires V4, où sont traités les couleurs et les formes, et V5, où sont traités la vitesse et la direction des objets pour en percevoir le mouvement. Enfin

2.1 Qu'est ce que l'attractivité faciale? . . . . .	5
2.1.1 Perception visuelle . . . . .	5
2.1.2 Mécanismes de l'attractivité des visages . . . . .	6
2.2 Perception et beauté . . . . .	8
2.2.1 Généralités . . . . .	8
2.2.2 Fluence et prototypicalité . . . . .	9
2.3 Sélection sexuelle et choix de partenaire . . . . .	10
2.3.1 Définitions et généralités	10
2.3.2 Dimorphisme sexuel et impacts . . . . .	11

[1] : MORRIS (2012), « Animal Eyes (Oxford Animal Biology Series) – By Michael F. Land & Dan-Eric Nilsson »

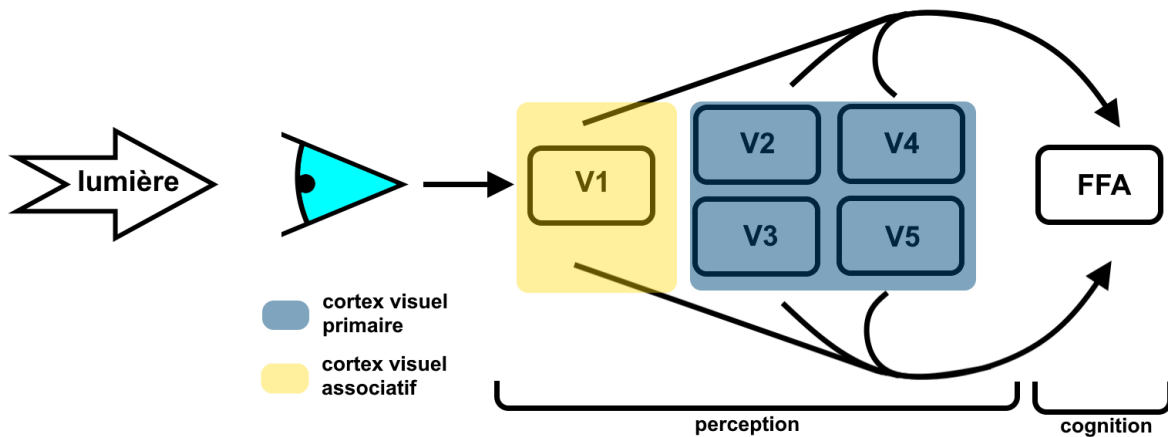


FIG. 2.1 : Représentation schématique du traitement de l'information des visages par le système visuel des mammifères, depuis l'acquisition de l'information lumineuse jusqu'au traitement perceptif et cognitif. Cette figure n'a pas vocation à représenter de manière exhaustive la structure détaillée du système visuel mais plutôt de présenter une vue globale des étapes du traitement visuel des visages (*production personnelle*)

le cortex Inféro-Temporal (IT) est impliqué dans la reconnaissance des objets et des formes à partir des informations des aires précédentes.

Une autre partie du cerveau, l'aire fusiforme des visages (FFA, pour Fusiform Face Area), est spécialisée dans la perception, la reconnaissance et la différenciation des visages. Elle est située dans le lobe temporal du cerveau, à la base du crâne. Cette partie a la particularité de prendre en considération l'ensemble des caractéristiques d'un visage de manière unifiée, contrairement à d'autres régions du cerveau qui traitent les caractéristiques des objets individuellement [2]. Ce sont principalement les informations de formes et de couleurs de l'aire V4 qui sont utilisées par la FFA pour reconnaître les visages. Au-delà du processus de perception, la FFA interagit avec des régions cérébrales impliquées dans des fonctions cognitives, en particulier la mémoire pour associer les visages perçus à des identités connues (Cf. Fig. 2.1).

[2] : KANWISHER et al. (2006), « The fusiform face area »

### 2.1.2 Mécanismes de l'attractivité des visages

On propose comme définition, volontairement générale et parcimonieuse, de l'attractivité la capacité d'un objet à attirer, à susciter l'intérêt. La notion d'objet est ici utilisée dans son sens phénoménologique, c'est à dire qu'il s'agit d'une entité ou d'un phénomène se présentant à l'expérience d'un sujet qui le perçoit. Une personne, a fortiori un visage, est donc un objet. Le concept de beauté, intrinsèquement lié à l'attractivité mais plus spécifique, sera développé dans la partie suivante.

Historiquement, les premiers travaux et réflexions autour de l'attractivité ont concerné les caractéristiques qui rendaient, au-delà des visages, des objets beaux universellement, comme le nombre d'or, la symétrie, ou le caractère fractal. Bien qu'incomplète, cette approche a connu des développements jusqu'à aujourd'hui. En vision par ordinateur, des travaux ont montré que des combinaisons de caractéristiques de bas niveau comme la luminance, le flou, le contraste, les couleurs ou encore la saturation [3, 4] peuvent prédire l'attractivité d'images. La symétrie est un critère considéré comme particulièrement attractif [5], ainsi que le caractère fractal, c'est-à-dire l'invariance à l'échelle [6]. Plus généralement, ces

[3] : TONG et al. (2004), *Classification of Digital Photos Taken by Photographers or Home Users*

[4] : DATTA et al. (2006), « Studying Aesthetics in Photographic Images Using a Computational Approach »

[5] : GRAMMER et al. (1994), « Human (Homo sapiens) facial attractiveness and sexual selection »

[6] : AMIRSHAHI et al. (2014), « Evaluating the Rule of Thirds in Photographs and Paintings »

critères objectifs et mesurables ont fait l'objet de nombreux travaux [7, 8].

Concernant les visages en particulier, la symétrie peut être un indicateur de qualité, comme par exemple dans le cas d'un individu qui aurait réussi à conserver un visage symétrique malgré des pressions environnementales [9]. En particulier, cela peut être un indicateur de bonne santé : les individus dont les visages sont plus symétriques ont moins de maladies respiratoires [10], ou un indicateur d'une meilleure fécondité [11, 12]. La centralité d'un visage, c'est-à-dire à quel point le visage ressemble à la majorité (*averageness* en anglais), est aussi un critère d'attractivité. Il pourrait s'agir d'un indicateur de résistance aux parasites car elle impliquerait des gènes plus variés donc des protéines moins communes auxquelles les pathogènes seraient moins adaptés [13]. Du moins, une préférence a été montrée pour des visages plus centraux [14]. D'autres caractéristiques sont particulièrement attractives comme la couleur rouge de certaines parties de la tête chez certains poissons [15], oiseaux [16] et primates non humains [17]. Des expressions faciales associées à des traits de personnalité sont aussi perçues comme attractives, comme des visages avenants, car ils signalent des capacités de coopération et d'investissement parental [18, 19]. Chez les humains, des traits comme la pilosité faciale chez les hommes [20] ou le maquillage chez les femmes [21] sont aussi des facteurs d'attractivité. Enfin, une attractivité plus importante a été relevée quand le visage observé ressemble à celui de l'observateur [22, 23]

Néanmoins, les préférences pouvant varier entre différents observateurs ou au cours de la vie d'un même observateur, l'aspect subjectif de celles-ci a été, plus tard, considéré. Au-delà des visages, l'appréciation d'œuvres d'arts est modulée selon le contexte artistique et historique, mais aussi selon les paramètres très subjectifs de l'expertise dans le domaine artistique de l'observateur [24] ainsi que ses émotions au moment de l'observation [25, 26].

Pour ce qui est des visages, des caractéristiques intrinsèques des observateurs peuvent aussi influencer leur attractivité, soit directement, soit en modifiant l'intensité de l'attractivité d'autres facteurs. Pour un observateur, la perception de sa propre attractivité va augmenter la force de l'attractivité de la symétrie des visages des autres personnes, de leur bonne santé apparente ou encore de la préférence, pour les femmes, pour les visages les plus masculins [27]. Certaines hormones ont aussi des effets similaires : lors du pic de fertilité du cycle menstruel, quand le taux d'oestrogène est le plus élevé, les femmes ont une préférence plus marquée pour les visages plus masculins [28-30]. Similairement, les hommes avec un taux de testostérone plus élevé ont une préférence pour les visages les plus féminins [31]. Pour les femmes, les visages les plus masculins sont considérés comme plus attirants dans l'hypothèse d'une relation courte qu'une relation longue [30, 32]. Par ailleurs, les femmes ont une aversion plus prononcée pour les visages présentant des caractéristiques associées à une maladie, comme la pâleur, quand elles ont un taux de progestérone plus élevé [33]. Plus généralement, les visages similaires à celui de l'observateur pouvant laisser présager un apparentement sont préférés [34], mais pas dans un contexte sexuel, pour éviter la consanguinité. Les caractéristiques de l'environnement vont aussi avoir un impact sur les préférences : quand l'environnement est moins favorable

[7] : BRACHMANN et al. (2017), « Computational and Experimental Approaches to Visual Aesthetics »

[8] : BALIETTI (2020), « The human quest for discovering mathematical beauty in the arts »

[9] : MØLLER (1997), « Developmental stability and fitness »

[10] : THORNHILL et al. (2006), « Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women »

[11] : MANNING et al. (1998), « Developmental Stability, Ejaculate Size, and Sperm Quality in Men »

[12] : MANNING et al. (1997), « Breast asymmetry and phenotypic quality in women »

[13] : THORNHILL et al. (1993), « Human facial beauty »

[14] : APICELLA et al. (2007), « Facial averageness and attractiveness in an isolated population of hunter-gatherers »

[15] : MILINSKI et al. (1990), « Female sticklebacks use male coloration in mate choice and hence avoid parasitized males »

[16] : PRYKE et al. (2005), « Red dominates black »

[17] : SETCHELL et al. (2005), « Dominance, Status Signals and Coloration in Male Mandrills (*Mandrillus sphinx*) »

[18] : OTTA et al. (1996), « Reading a smiling face »

[19] : HASSIN et al. (2000), « Facing faces »

[20] : NEAVE et al. (2008), « The effects of facial hair manipulation on female perceptions of attractiveness, masculinity, and dominance in male faces »

[21] : OSBORN (1996), « Beauty is as Beauty Does? »

[22] : DEBRUINE (2004), « Facial resemblance increases the attractiveness of same-sex faces more than other-sex faces »

[23] : DEBRUINE (2005), « Trustworthy but not lust-worthy »

[24] : DANTO (), *The Transfiguration of the Commonplace*

[25] : LEDER et al. (2004), « A model of aesthetic appreciation and aesthetic judgments »

[26] : SILVIA (2005), « Emotional Responses to Art »

[27] : LITTLE et al. (2001), « Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. »



[28] : JOHNSTON et al. (2001), « Male facial attractiveness »

[29] : JONES et al. (2005), « Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high »

[30] : PENTON-VOAK et al. (1999), « Menstrual cycle alters face preference »

[31] : WELLING et al. (2008), « Men report stronger attraction to femininity in women's faces when their testosterone levels are high »

[30] : PENTON-VOAK et al. (1999), « Menstrual cycle alters face preference »

[32] : LITTLE et al. (2002), « Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. »

[33] : RHODES et al. (2003), « Does sexual dimorphism in human faces signal health? »

[34] : HAMILTON (1964), « The genetical evolution of social behaviour. I »

[35] : DEBRUINE et al. (2010), « The health of a nation predicts their mate preferences »

[36] : BROOKS et al. (2011), « National income inequality predicts women's preferences for masculinized faces better than health does »

[37] : BORNSTEIN (1989), « Exposure and Affect »

[38] : ZAJONC (2001), « Mere Exposure »

[39] : ZAJONC et al. (1969), « Exposure and affect »

[40] : LITTLE et al. (2011), « Social learning and human mate preferences »

[41] : AUGUSTIN et al. (2012), « All is beautiful? »

[42] : BUSS et al. (1986), « Preferences in Human Mate Selection »

[43] : ELDER (1969), « Appearance and education in marriage mobility »

[44] : CASH et al. (1985), « The Eye of the Beholder »

[45] : RIGGIO et al. (1984), « The Role of Nonverbal Cues and Physical Attractiveness in the Selection of Dating Partners »

[46] : BERSCHIED et al. (1971), « Physical attractiveness and dating choice »

[47] : WALSTER et al. (1966), « Importance of physical attractiveness in dating behavior. »

[48] : SIGALL et al. (1975), « Beautiful but Dangerous »

[49] : LANGLOIS et al. (2000), « Maxims or myths of beauty? »

[50] : DION et al. (1972), « What is beautiful is good. »

à une bonne santé, c'est-à-dire quand il y a une prévalence de pathogènes plus importante et un accès plus difficile aux soins, les femmes préfèrent les visages les plus masculins [35]. C'est aussi le cas quand il y a plus d'actes violents dans l'environnement[36]. L'exposition répétée, créant une impression de familiarité, rend aussi les visages plus attractifs[37-39]. Enfin, le contexte social a une influence dans la mesure où les visages connus comme étant préférés par d'autres sont considérés comme plus attractifs[40].

Tous ces critères montrent que l'attractivité faciale revêt de multiples facettes en lien avec l'esthétique et la beauté. Les nuances entre ces deux termes sont détaillées en Encadré 1 [[41]]. Nous allons donc par la suite nous intéresser à cet aspect spécifique de l'attractivité : la beauté.

### Encadré 1 : Beauté et esthétique de quoi parle t-on exactement ?

Les concepts de beauté et d'esthétique sont étroitement liés, mais des travaux ont insisté sur la nécessité de les différencier. La beauté désigne spécifiquement **une expérience agréable, hédonique** alors que l'expérience esthétique est plus générale et peut inclure des expériences mixtes ou désagréables comme le vertige. De fait, cette expérience esthétique, au-delà du plaisir, est basée sur la notion d'attention, d'intérêt. Elle sera d'autant plus intense quand il y aura un bon compromis entre plaisir de l'expérience et intérêt.

Dans nos travaux, la beauté décrit spécifiquement le plaisir issu de la facilité de traitement de l'information uniquement pendant le traitement ascendant de l'information.

## 2.2 Perception et beauté

Il est difficile d'aborder l'attractivité d'un élément visuel, ici les visages, sans traiter le concept de beauté, auquel elle est intrinsèquement liée. Darwin avait défini la beauté comme le plaisir procuré par certaines couleurs, formes, et sons. Ainsi, nous verrons dans cette partie que définir, comprendre et même prédire la beauté est une question qui préoccupe les scientifiques, mais aussi les philosophes et les artistes depuis l'aube de l'humanité.

### 2.2.1 Généralités

La beauté confère de nombreux avantages sociaux et sexuels. Les humains, autant les hommes que les femmes, accordent de l'importance à la beauté de leurs partenaires potentiels [42]. Dans les sociétés humaines, être beau permet d'avoir un meilleur salaire [43], d'être embauché plus facilement[44], d'avoir plus de rendez-vous amoureux[45-47], d'avoir une peine plus clémentine au tribunal[48] ainsi que beaucoup d'autres stéréotypes ayant pour points communs l'idée que ce qui est beau est bon [49, 50].

La beauté est un phénomène complexe qui implique la perception, la cognition et les émotions [51-53]. L'implication de la perception s'explique par l'hypothèse que la beauté émerge d'une activité neuronale

particulière du système visuel (décrit en partie 2.1.1) [53-55], réagissant à des caractéristiques intrinsèques aux objets. L'aspect cognitif s'explique par l'importance du contexte et des différences interindividuelles, comme développé en partie 2.1.2. Déjà en 1757 dans *De la norme du goût*, Hume affirmait que "la beauté n'est pas une qualité inhérente aux choses elles-mêmes, elle existe seulement dans l'esprit qui la contemple et chaque esprit perçoit une beauté différente". Au XXe siècle, les deux aspects de l'attractivité développés dans la partie précédente, objectif et subjectif, ont été pris en compte simultanément à travers l'idée que la beauté émerge non pas seulement des caractéristiques de l'objet ou de celles de l'observateur, mais de l'interaction entre les deux[56]. Nous allons nous attarder dans la partie suivante sur une manière de formaliser cette interaction : la fluence.

### 2.2.2 Fluence et prototypicalité

La théorie de la fluence, développée en psychologie, concerne la manière dont le cerveau traite l'information[57]. La fluence est définie comme une sensation de facilité dans le traitement de l'information, provoquant une expérience agréable qui correspond à la beauté. Elle permet de faire le lien entre les propriétés de l'objet perçu et le système perceptif et cognitif qui le traite et correspond donc à l'idée d'interaction objet-sujet précédemment décrite. De plus, elle explique les préférences pour certaines caractéristiques évoquées précédemment : par exemple, la symétrie et le caractère fractal impliquent une redondance de l'information la rendant plus facile à traiter.

Elle peut être mesurée de différentes manières[58]. Des mesures indirectes existent, comme le temps de réaction, mais c'est une mesure limitée car des facteurs de confusion liés au système moteur la rende peu fiable. La manière probablement la plus précise de la mesurer impliquerait une analyse de connectivité fonctionnelle basée sur l'imagerie cérébrale, ce qui n'a pas encore été réalisé. D'autres mesures indirectes comme la prototypicalité ont été explorées. Un prototype est un représentant moyen d'une catégorie [59]. Expliqué différemment, c'est l'image mentale la plus probable que l'on peut se faire d'un objet quand nous y pensons (Cf. Tab. 2.1). Les prototypes ont tendance à être préférés et nous pouvons relier cette préférence à l'attractivité de la centralité (*averageness*) détaillées en partie 2.1.2. Ces prototypes sont préférés car ils sont faciles à catégoriser, à classifier et donc à reconnaître. Cette facilité accrue dans le traitement de l'information pourrait donc expliquer que les prototypes sont, d'une manière générale, jugés comme attractifs. Nous voyons alors le lien avec la fluence puisque ce codage efficient implique une facilité du traitement de l'information.

Nous avons dans les parties précédentes évoqué la masculinité et la féminité comme des critères d'attractivité parmi d'autres. Si nous définissons ces deux notions comme la proximité respective à un prototype masculin ou féminin, nous voyons que ce type de préférence peut également s'expliquer par la théorie de la fluence. Pour autant, les caractéristiques et la perception de la masculinité et de la féminité sont liées à l'évolution des comportements et des signaux sociaux-sexuels que nous allons détailler dans la partie suivante.

[51] : JACOBSEN (2006), « Bridging the Arts and Sciences »

[52] : MARKOVIĆ (2012), « Components of aesthetic experience »

[53] : REDIES (2015), « Combining universal beauty and cultural context in a unifying model of visual aesthetic experience »

[53] : REDIES (2015), « Combining universal beauty and cultural context in a unifying model of visual aesthetic experience »

[54] : TAYLOR et al. (2005), « Fractals »

[55] : REDIES et al. (2007), « Fractal-like image statistics in visual art »




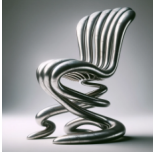

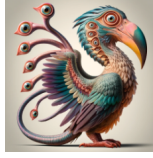
[56] : ISHIZU et al. (2011), « Toward A Brain-Based Theory of Beauty »

[57] : REBER et al. (2004), « Processing Fluency and Aesthetic Pleasure »

[58] : OPPENHEIMER (2008), « The secret life of fluency »

[59] : VOGEL et al. (2021), « The architecture of prototype preferences »

**TAB. 2.1 : Exemples d'objets de la vie quotidienne correspondant ou non à des prototypes.** Pour chacun d'entre eux, on a utilisé le modèle génératif d'images à partir de texte Dall-e avec comme prompt : "crée une image d'un objet normal" pour les prototypes, et "crée une image d'un objet bizarre" pour les non prototypes, en remplaçant "objet" par "chaise", "arbre", ou "oiseau".

	chaise	arbre	oiseau
prototype			
non prototype			

## 2.3 Sélection sexuelle et choix de partenaire



**FIG. 2.2 : Un paon bleu mâle adulte faisant la roue** (image libre de droits)



**FIG. 2.3 : Un cerf mâle** (image libre de droits)



**FIG. 2.4 : Un cardinal rouge** (image libre de droits)

Les espèces évoluent. Autrement dit, à travers les générations, les phénotypes des organismes, les caractéristiques des populations, mais surtout les gènes, changent à travers les générations. Le mécanisme majeur de cette évolution est la sélection naturelle : les caractéristiques les plus avantageuses dans un environnement vont augmenter la survie des individus les possédant, et donc leur descendance. Cependant, ce mécanisme n'explique pas toutes les caractéristiques des êtres vivants. Certains caractères extravagants comme la queue des paons (Cf. Fig. 2.2) ou les bois des cervidés (Cf. Fig. 2.3) sont encombrants pour le déplacement et très visibles pour les prédateurs. Néanmoins, ils ne présentent pas d'intérêt pour la survie des individus. Cela s'explique par une sélection de certains traits (Cf. encadré 2) sur un avantage autre que la survie : l'accès aux partenaires sexuels. Il s'agit de sélection sexuelle. Nous allons expliquer de quoi il s'agit et les liens avec le sujet de la thèse en partie 2.3.1, puis nous focaliser sur un aspect particulier : le dimorphisme sexuel, en partie 2.3.2.

### 2.3.1 Définitions et généralités

Plus rigoureusement, la sélection sexuelle est définie comme "le processus par lequel certains caractères sont sélectionnés en vertu de l'avantage qu'ils confèrent dans l'accès aux partenaires sexuels" [60]. Elle se divise en deux catégories. La sélection intrasexuelle, correspond aux traits sélectionnés pour obtenir un avantage compétitif entre individus du même sexe, par exemple lors de combats (comme les bois des cerfs). La sélection intersexuelle correspond aux traits sélectionnés pour obtenir un avantage pour être choisi par un partenaire du sexe opposé (Mate choice dans la littérature [61]). Par exemple, le plumage rouge vif du cardinal rouge mâle *Cardinalis cardinalis* est utilisé pour attirer les femelles (Cf. Fig. 2.4).

Les traits sélectionnés dans ce contexte sont les caractères sexuels secondaires. Les caractères sexuels primaires correspondent aux traits directement impliqués dans la reproduction : les organes génitaux, permettant la production de gamètes, l'accouplement et la gestation. Les caractères sexuels secondaires sont par opposition ceux qui ne sont pas directement impliqués dans le processus de reproduction mais qui jouent un rôle dans

[60] : CÉZILLY et al. (2010), « La sélection sexuelle »

[61] : ROSENTHAL (2017), *Mate Choice*

l'attraction des partenaires et la compétition entre les individus du même sexe, donc dans la sélection sexuelle.

### Encadré 2 : Traits signaux et préférences

Chez les êtres vivants, quand on parle de **traits**, on s'intéresse à des caractères phénotypiques, qui peuvent être observés.

Ainsi, un **signal** visuel peut être un trait. Il s'agit d'une information visuelle du phénotype d'un individu permettant aux autres individus d'en savoir plus.

Une **préférence** est aussi un trait, y compris une préférence pour un signal.

Les traits des visages en particulier sont considérés comme des signaux honnêtes, qui n'émettent pas de fausse information permettant à l'individu de "tricher" vis-à-vis de ses congénères [62].

[62] : GANGESTAD et al. (2005), « The Evolution of Human Physical Attractiveness »

La sélection intersexuelle nécessite une préférence chez l'un des partenaires, qui peut avoir deux origines. Dans un premier cas, une caractéristique phénotypique visuelle peut apparaître en premier lieu chez l'un des deux sexes, puis une préférence pour ce signal sexuel spécifiquement apparaît secondairement chez le sexe opposé. Dans le second cas, une préférence peut être dite latente, c'est-à-dire exister alors que le signal sexuel n'existe pas encore, et ne s'exprimer que si le signal apparaît secondairement. Ces préférences ne naissent donc non pas d'une sélection sexuelle mais sont généralement liées à d'autres fonctions (comme l'alimentation) et sont à l'origine soumises à sélection naturelle.

### 2.3.2 Dimorphisme sexuel et impacts

Le dimorphisme sexuel est la différence entre les mâles et les femelles d'une même espèce. Cette différence peut impliquer la taille et la morphologie, mais aussi la couleur ou les comportements. Comme vu précédemment, le dimorphisme sexuel peut provenir d'une pression de sélection sexuelle aboutissant à des traits exagérés chez l'un ou l'autre des sexes d'une même espèce. Toutefois, la sélection sexuelle n'est pas la seule explication de ce dimorphisme : certaines caractéristiques dimorphiques peuvent s'expliquer par le rôle différent dans la reproduction, la gestation et les soins aux jeunes, ou la recherche de nourriture. Le terme de dimorphisme sexuel peut à la fois désigner la différence entre les deux sexes d'une espèce en général, mais aussi, pour un individu en particulier, la différence d'avec l'autre sexe.

Le régime d'appariement des espèces a une influence sur l'intensité du dimorphisme sexuel. Dans des espèces monogames, le dimorphisme sexuel est plus faible car il y a moins de pression pour un sexe de développer des traits exagérés pour attirer l'autre sexe. À l'inverse, dans les régimes de polygynie (un mâle s'accouple avec plusieurs femelles) et polyandrie (une femelle s'accouple avec plusieurs mâles), le dimorphisme est plus important, car la compétition accrue induite par ce déséquilibre en terme de nombre de partenaires favorise le développement de traits secondaires exagérés.



# L'intelligence artificielle pour modéliser l'attractivité faciale

# 3

On a vu jusqu'à maintenant que l'attractivité des visages est multifactorielle. Plus encore, ces facteurs sont liés entre eux, et s'inscrivent dans différents contextes : organisation sociale, fonctionnement du cerveau et évolution des espèces. Cette complexité caractérisant l'attractivité faciale rend nécessaire l'utilisation d'outils puissants pour la comprendre et définir clairement le rôle des facteurs qui la régissent et leurs interactions.

Dans cette partie, nous proposons l'intelligence artificielle (IA) comme un outil de modélisation de l'attractivité des visages. L'image numérique est le meilleur moyen de transformer un visage en structure de données compréhensible par un ordinateur. Cependant, les images présentent une forte autocorrélation spatiale (les pixels proches se ressemblent) et sont des structures de données avec un nombre de variables très élevé (il y a des centaines de milliers de pixels, de chacun trois variables, dans une image). Ces contraintes sont difficiles à gérer pour les modèles de statistiques traditionnels, mais plus en adéquation avec les forces des modèles d'apprentissage profond qui ont émergé depuis une dizaine d'années. Ces modèles sont en effet capables de réduire la dimensionnalité des images tout en captant une grande partie de la variance de ces données et des facteurs qui les structurent. De plus, dans les espaces algébriques créés par ces modèles, les variables en jeu peuvent être démêlées (*disentanglement*<sup>1</sup>), les rendant plus faciles à expliquer et à interpréter. En outre, bien que les modèles d'IA n'aient pas été conçus comme des copies du cerveau humain, la manière dont ils traitent l'information peut, dans une certaine mesure, être assimilée au traitement des images par le cortex visuel que nous avons détaillé en partie 2.1.1. Enfin, des bases de données de visages annotés, indispensable à l'entraînement de tels modèles, accessible à la communauté scientifique, existent autant chez les humains [63] que chez certains primates [64].

Nous commencerons par définir et expliquer des concepts et notions d'intelligence artificielle en partie 3.1. Ensuite, en partie 3.2, nous argumenterons sur la nécessité de disposer de modèles et de métriques pertinents pour modéliser l'attractivité des visages.

<b>3.1 Intelligence artificielle</b>	<b>14</b>
3.1.1 Apprentissage automatique . . . . .	14
3.1.2 Réseaux de neurones artificiels et apprentissage profond . . . . .	16
3.1.3 Intelligence artificielle générative . . . . .	20
<b>3.2 Un modèle et des métriques</b> . . . . .	<b>22</b>

1 : Dans cette partie, nous donnerons en italique les traductions en anglais de certains termes, car celles-ci sont souvent plus utilisées que les équivalents en français.

[63] : LIANG et al. (2018), « SCUT-FBP5500 »

[64] : TIEO et al. (2023), « The Mandrillus Face Database »

### 3.1 Intelligence artificielle

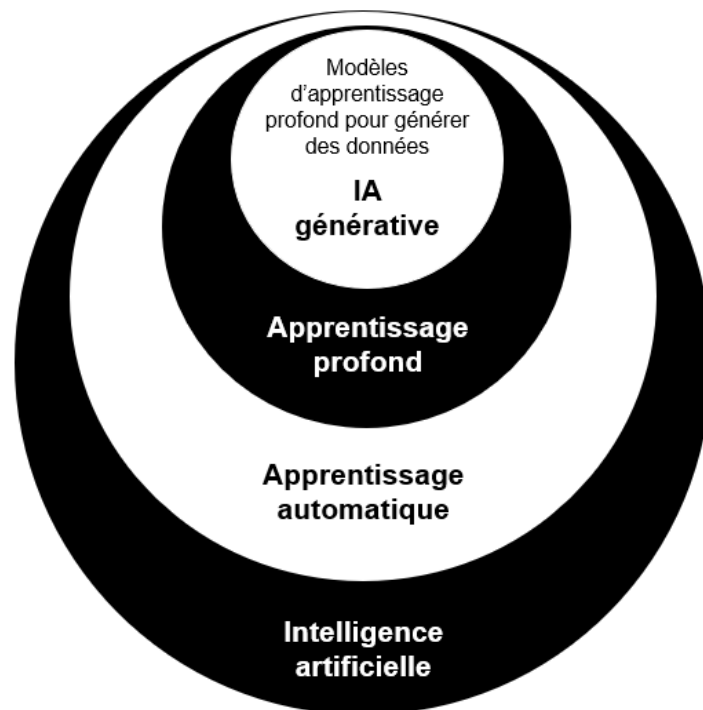


FIG. 3.1 : Diagramme de l'imbrication de certaines catégories d'intelligence artificielle (production personnelle)

L'intelligence artificielle (IA) est un terme général et dont la définition est peu consensuelle. C'est surtout le terme intelligence qui fait débat, mais qui peut être exprimé comme la capacité d'apprendre d'une expérience, de s'adapter à de nouvelles situations ou de résoudre de nouveaux problèmes. Un des premiers types d'IA apparu dans l'histoire, l'IA dite symbolique repose sur la manipulation de symboles et de règles explicites pour représenter des connaissances et résoudre des problèmes. Cette approche est néanmoins rigide car elle dépend de règles explicites fixées par des humains. Une approche plus récente, appelée IA statistique ou IA neuronale, correspond plus à l'automatisation de ce comportement intelligent [65]. Toutefois, des questions subsistent : est ce que l'intelligence est une seule capacité, ou la somme de plusieurs facultés distinctes ? Est ce qu'elle existe a priori ou est ce qu'elle peut être apprise ? Quelle est la place de la créativité et de l'intuition dans l'intelligence ? Est ce qu'elle concerne uniquement les humains ? Les animaux ? Les êtres vivants ? [65]. Malgré cette incertitude, l'intelligence artificielle a l'avantage d'être un terme général et englobant, que tout le monde a déjà entendu, et c'est pourquoi nous avons choisi de l'intégrer dans le titre de cette thèse. Dans cette partie, nous expliquerons des partitions de plus en plus spécifiques (mais non exhaustives) de l'IA (Cf. Fig. 3.1) : l'apprentissage automatique en partie 3.1.1, l'apprentissage profond en partie 3.1.2 et l'IA générative en partie 3.1.3. .

[65] : CHOWDHARY (2020), *Fundamentals of Artificial Intelligence*

[65] : CHOWDHARY (2020), *Fundamentals of Artificial Intelligence*

#### 3.1.1 Apprentissage automatique

L'apprentissage automatique (*machine learning*), bien qu'étant une sous-catégorie de l'IA, est une discipline plus précisément définie et segmen-

tée : elle consiste à programmer des ordinateurs pour qu'ils apprennent à partir de données [66]. Pour exécuter une tâche via le système d'un ordinateur, il faut écrire un algorithme : une séquence d'instructions qui transforme des données d'entrée (*input*) en données de sortie (*output*). C'est facile pour certaines tâches, et plus dur pour d'autres. Par exemple, alors que les humains peuvent facilement reconnaître un visage sur un portrait ou conduire une voiture, la tâche est moins triviale pour un ordinateur. L'apprentissage automatique permet d'apprendre à faire ce type de tâches. L'idée est de commencer avec un modèle comportant beaucoup de paramètres ajustables pour transformer des entrées en sortie. L'apprentissage du modèle consistera alors à ajuster tous ces paramètres pour que la tâche soit correctement effectuée [67]. Le modèle final sera alors l'approximation d'une fonction reliant les données d'entrée et les données de sortie.

On divise l'apprentissage automatique en deux grandes approches : l'apprentissage supervisé, et l'apprentissage non supervisé. En apprentissage automatique supervisé, le modèle est entraîné sur des données d'entrées associées à des étiquettes (*labels*) correspondant aux données de sortie. L'apprentissage du modèle consiste donc à prédire les labels déjà connus dans un premier temps, pour ensuite être capable de fonctionner sur des données non étiquetées. Les régressions linéaires, forêts aléatoires (*random forest*) et machines à vecteur de support (*Support Vector Machine, SVM*) sont des types d'apprentissage supervisé. À l'inverse, l'apprentissage non supervisé est une approche où le modèle est entraîné sur des données sans étiquette. L'objectif est de découvrir des structures cachées, des motifs, ou des groupes dans les données. Les k-moyennes (*k-means*) et les Analyses en Composantes Principales (ACP) sont des types d'apprentissage non supervisé [66, 67]. Le concept de *cluster* (la traduction française, grappe, est désuète), est fondamental dans ces approches (ou en aval de ces approches) et correspond à des ensembles de données regroupant des caractéristiques communes.

D'autres approches existent comme l'apprentissage par renforcement (*reinforcement learning*) où un agent apprend à prendre des décisions en interagissant avec un environnement et en recevant des récompenses ou des punitions tout en cherchant à maximiser la récompense totale au fil du temps. Nous mentionnons aussi l'apprentissage semi-supervisé, hybride des apprentissages supervisé et non supervisé, qui utilise à la fois des données étiquetées et non étiquetées pour entraîner des modèles [66, 67].

Il existe plusieurs types de tâches en apprentissage automatique. La classification a pour objectif de prédire la catégorie à laquelle appartient une nouvelle donnée, parmi des classes déjà connues. Il s'agit d'une approche supervisée, à ne pas confondre avec le clustering, tâche non supervisée consistant à regrouper des données similaires ensemble, mais sans classes prédéterminées. La régression, approche classiquement utilisée en statistiques, permet de prédire des valeurs continues. La réduction de dimensionnalité permet de réduire un nombre de variables d'entrée en un plus petit nombre de variables de sortie [66, 67].

Pour fabriquer un modèle d'apprentissage automatique performant, il faut l'entraîner. Les données qui vont servir à l'entraînement sont généralement divisées en 3 parties : l'ensemble d'entraînement (*train set*) va permettre au modèle d'apprendre les relations et les motifs (*patterns*)

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*



[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[67] : ALPAYDIN (2020), *Introduction to Machine Learning, fourth edition*

dans les données. L'ensemble de validation (*validation set*) va permettre d'évaluer les performances du modèle pendant l'entraînement pour ajuster ses paramètres. L'ensemble de test (*test set*) va permettre d'évaluer les performances du modèle une fois que l'entraînement est terminé. Pour améliorer la robustesse de ce processus, l'entraînement du modèle peut être réitéré de sorte que la division entre ensemble d'entraînement et ensemble de validation soit différente à chaque itération : cette technique s'appelle la validation croisée (*cross validation*). [66, 67]

La gestion du surapprentissage est un des enjeux majeurs du machine learning. Il s'agit d'une situation où le modèle est trop spécialisé sur les données d'entraînement, et n'est pas capable de généraliser sur de nouvelles données similaires. L'utilisation de jeux de validation et de test, ainsi que de validation croisée peut permettre de réduire ce problème [66, 67].

En apprentissage automatique, les données doivent être décrites de manière structurée et compréhensible pour être utilisées par des algorithmes via le système d'un ordinateur. Il peut s'agir de chaîne de caractères, de tableaux, de séries temporelles ou encore d'images. Concrètement, il s'agit de vecteurs ou de matrices, de plus ou moins grande dimension. Par exemple, une image est une matrice de pixels, en 3 dimensions : hauteur, largeur et profondeur (car un pixel se décompose en plusieurs canaux, par exemple 3 dans une image classique RGB (*Red Green Blue*)). Pour autant, on peut aussi voir ces données comme des objets (points, vecteurs, nuages) dans des espaces géométriques de grande dimension.

Malgré leur puissance, beaucoup d'algorithmes traditionnels d'apprentissage automatique peuvent être limités sur des tâches vraiment complexes. Sur des données non structurées, où les caractéristiques ne sont pas précisément annotées comme des images, des vidéos, du texte ou des sons, ils ne parviennent pas, ou pas aussi bien qu'un humain, à résoudre efficacement des problèmes comme reconnaître une personne ou une voix, ou traduire d'un langage à un autre. Toutefois, une classe particulière de l'apprentissage automatique, l'apprentissage profond (*deep learning*), permet de pallier à des limites, et de manière de plus en plus performantes ces dernières années.

### 3.1.2 Réseaux de neurones artificiels et apprentissage profond

Les réseaux de neurones artificiels sont des modèles d'apprentissage automatique, composés de sous-unités appelées neurones artificiels. Ces composants de base sont reliés entre eux, chacun recevant en entrée les sorties des neurones précédents, et produisant une sortie vers les neurones suivants. Les neurones sont organisés en couches successives. La première couche du réseau reçoit les données d'entrée du modèle et la dernière produit la sortie du modèle. Entre les deux, il y a plusieurs couches, appelées couches cachées (*hidden layers*) car elles ne correspondent pas directement aux données d'entrée ou de sortie mais à des transformations intermédiaires plus abstraites (Cf. Fig. 3.2) [68]. L'architecture d'un réseau désigne la structure et l'organisation des composants du réseau comme le nombre de couches, le nombre de neurones par couche, et la manière dont les neurones sont connectés entre eux.

[68] : GOODFELLOW et al. (2018), *L'apprentissage profond / Ian Goodfellow, Yoshua Bengio, Aaron Courville*

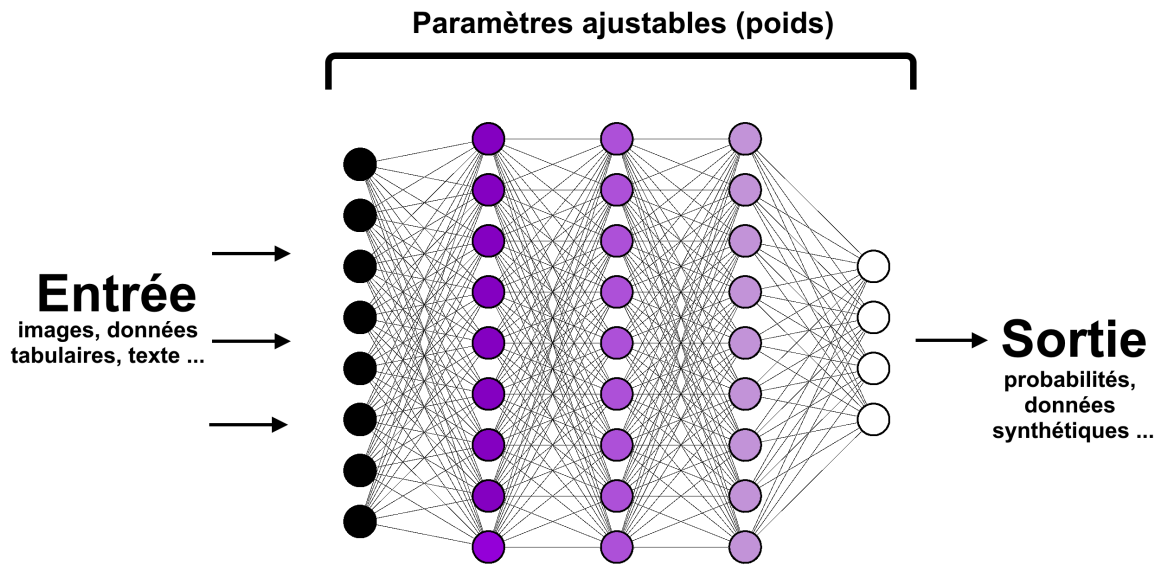


FIG. 3.2 : Cas général d'un réseau de neurones artificiel En noir la couche d'entrée, en violet les couches cachées, en blanc une couche de sortie (*production personnelle*)

L'apprentissage profond (Deep learning) correspond à des réseaux de neurones avec beaucoup de couches cachées (Cf. Fig. 3.2). Cela permet au réseau d'apprendre des représentations hiérarchiques des données, capturant des motifs de plus en plus complexes à chaque couche successive [66, 68], jusqu'à obtenir un vecteur, appelé vecteur latent ou vecteur caractéristique. Ces modèles permettent de gérer des données plus complexes que les approches précédemment évoquées (Cf. Encadré 3 [69, 70]).

### Encadré 3 : Science des données et démêlage (*disentanglement*)

Les méthodes linéaires de réduction de dimensionnalité, comme l'Analyse en Composantes Principales (ACP), sont souvent inefficaces pour des données qui possèdent une structure non linéaire complexe. Dans le cas du Swiss Roll (Cf. Fig. 3.3), les données sont enroulées dans un espace tridimensionnel, et une projection linéaire ne peut pas capturer cette complexité. Les algorithmes de clustering comme K-means utilisent des distances euclidiennes pour partitionner les données. Cependant, sur le Swiss Roll, les points proches en termes de distance euclidienne peuvent être éloignés sur le motif sous-jacent et ne pas appartenir au même cluster.

Des méthodes comme l'Isomap, le t-SNE (*t-distributed Stochastic Neighbor Embedding*) ou encore l'UMAP (*Uniform Manifold Approximation and Projection*) sont conçues pour gérer des structures non linéaires complexes. Elles tentent de préserver la géométrie locale et la structure globale des données. En utilisant ces techniques, le Swiss Roll peut être déroulé en une forme plate, révélant ainsi la structure intrinsèque en deux dimensions des données. Cela permet une meilleure visualisation et compréhension.

Néanmoins, les formes que peuvent prendre les données dans des

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and TensorFlow*

[68] : GOODFELLOW et al. (2018), *L'apprentissage profond* / Ian Goodfellow, Yoshua Bengio, Aaron Courville

[69] : BRAHMA et al. (2016), « Why Deep Learning Works »

[70] : BURGOYNE et al. (2007), *Non-linear scaling techniques for uncovering the perceptual dimensions of timbre*

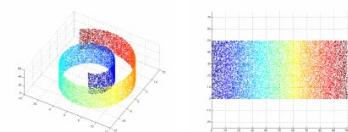


FIG. 3.3 : le "Swiss Roll", un exemple de données emmêlées de manière non linéaire À gauche, les données sont présentées dans leur forme originale. À droite, les données sont présentées telles qu'elles devraient être déroulées pour être interprétables.

2 : merci de ne pas faire cela avec ce manuscrit de thèse

[71] : ARBIB (2000), « Warren McCulloch's Search for the Logic of the Nervous System »

[72] : RUMELHART et al. (1986), « Learning representations by back-propagating errors »

[73] : GRAVES et al. (2006), *Connectionist temporal classification*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

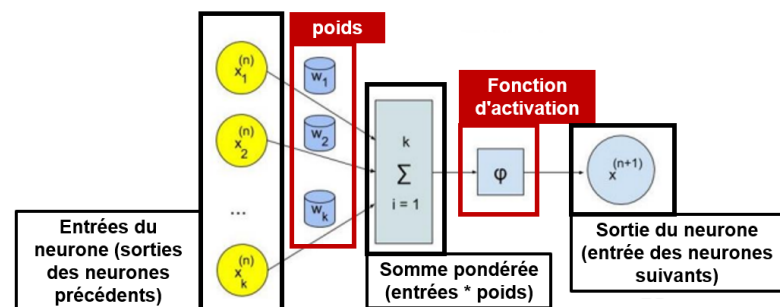
[68] : GOODFELLOW et al. (2018), *L'aprentissage profond / Ian Goodfellow, Yoshua Bengio, Aaron Courville*

FIG. 3.4 : Fonctionnement d'un neurone artificiel (production personnelle)

espaces à haute dimension peuvent être encore plus complexes (le terme *manifold* est parfois utilisé). Pour visualiser une telle complexité, on peut tracer deux points proches sur une feuille de papier, puis la chiffonner en une boule<sup>2</sup>. Les deux points peuvent se retrouver spatialement très proches ou très éloignés, de manière décorrélée de leur position relative initiale. Les méthodes mentionnées peuvent alors être limitées pour démêler la relation entre les deux points, d'autant plus que la dimensionnalité de l'espace est importante. L'apprentissage profond peut par contre permettre de pallier à cette limite en réussissant mieux à démêler ces données.

Historiquement, les modèles initiaux, très simples, basés sur des neurones artificielles, datent du milieu du 20e siècle [71], mais les premiers réseaux de neurones artificiels à proprement parler sont apparus dans les années 70 [72]. Enfin, les modèles d'apprentissage profond ont émergé plus récemment, dans les années 2010. Cela a été permis par des ordinateurs plus performants permettant de faire fonctionner ces modèles contenant un très grand nombre de paramètres. Ceux-ci sont dès lors devenus l'état de l'art de beaucoup de tâches d'apprentissage automatique [73].

Un neurone artificiel, aussi appelé neurone formel, est une unité prenant en entrée les sorties des neurones précédents. Chacune de ces entrées est pondérée, et les poids sont les paramètres variables qui sont optimisés lors de l'entraînement du réseau. La somme pondérée des entrées est ensuite transformée par une fonction d'activation. Il s'agit d'une fonction qui va introduire des relations non linéaires entre les différents neurones, afin d'étendre la capacité du modèle à capturer des relations complexes dans les données. Par exemple, une fonction d'activation répandue est la fonction ReLU (Rectified Linear Unit), qui renverra 0 si la somme pondérée est négative et la même valeur si elle est positive :  $ReLU(x) = \max(0, x)$ . Le résultat de la fonction d'activation est ensuite transmis aux neurones suivants (Cf. Fig. 3.4) [66, 68].



Pour ajuster les poids des neurones et ainsi obtenir les sorties souhaitées à partir des entrées, avec une erreur minimale, les réseaux de neurones suivent un processus d'entraînement comme expliqué en partie 3.1.1. Pour cela, l'algorithme de rétropropagation (*backpropagation*) est utilisé. Pendant l'entraînement, une erreur de prédiction est calculée avec une fonction dite de coût (*loss function*) entre la sortie du réseau et la valeur réelle attendue. Les poids de chaque neurone sont alors ajustés dans le sens inverse du réseau, selon leur contribution à l'erreur finale [66, 68].

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[68] : GOODFELLOW et al. (2018), *L'aprentissage profond / Ian Goodfellow, Yoshua Bengio, Aaron Courville*

Les poids initiaux avant l'entraînement d'un réseau sont aléatoires. Néanmoins, pour accélérer le processus, il est possible de partir des poids d'un

réseau d'architecture identique qui a été entraîné sur une tâche similaire. C'est l'apprentissage par transfert (*transfer learning*). Par exemple, pour entraîner un réseau à reconnaître des images de tournesols, nous pourrions commencer avec un réseau entraîné à reconnaître des images d'autres plantes car les caractéristiques apprises seraient proches [66, 68].

Nous avons jusqu'à présent expliqué le fonctionnement général des réseaux de neurones. Il existe des grandes familles de réseaux possédant chacune des caractéristiques propres en termes d'architecture les rendant performants sur des tâches précises. Nous présentons ici certaines de ces familles parmi les plus répandues.

Les réseaux de neurones convolutifs (CNNs - *Convolutional Neural Networks*) sont une classe de réseaux de neurones principalement utilisés pour l'analyse des données visuelles. Ils se distinguent par leur capacité à capturer les caractéristiques spatiales et les relations locales dans les images, ce qui les rend particulièrement efficaces pour les tâches de vision par ordinateur. Leur architecture se caractérise par la présence répétée de deux types de couches de neurones spécifiques. Les couches de convolution, qui appliquent des filtres spécialisés dans certains types de motifs (comme des textures ou des bords) pour extraire des caractéristiques locales. Formellement, un filtre est une matrice de poids qui correspondent aux poids des neurones décrits précédemment. Les couches de pooling réduisent la dimensionnalité des matrices (on parle alors de cartes de caractéristiques, ou *feature map*) issues des couches de convolution, en prenant généralement leur valeur maximale. L'intérêt de prendre le maximum plutôt que la moyenne est de créer une invariance à la déformation et à l'échelle des caractéristiques d'une image, pour mieux généraliser. Plus une couche de convolution est située au début du réseau, plus elle va extraire des caractéristiques simples et localisées comme des inclinaisons ou des contrastes. À l'inverse, plus elle est située à la fin du réseau, plus elle va extraire des caractéristiques abstraites et conceptuelles comme des objets ou des visages. Tout à la fin du réseau, les cartes de caractéristiques sont aplaties en un vecteur, qui est ensuite passé à travers une ou plusieurs couches où les neurones sont entièrement connectés. Cela permet de combiner les caractéristiques extraites pour effectuer des prédictions finales et obtenir la sortie du modèle (Cf. Fig. 3.5) [66, 68]. Le fonctionnement d'un CNN peut être assimilé au système visuel des mammifères [74] que nous avons décrit en partie 2.1.1, et ainsi peut le modéliser. Ce point sera développé en partie 3.2.

Les réseaux de neurones récurrents (RNNs, *Recurrent Neural Networks*) sont utilisés en traitement du langage naturel (NLP, *Natural Language Processing*) et en analyse de séries temporelles. Ils ont la particularité de posséder des connexions internes qui bouclent sur eux-mêmes, ce qui permet de capturer des dépendances temporelles dans les séquences de données. Les LSTMs (*Long Short-Term Memory*) sont une variante des RNN qui introduisent une cellule de mémoire capable de maintenir l'information sur de longues périodes. Les réseaux de neurones en graphes (GNN, *Graph Neural Networks*) permettent de traiter des données structurées en graphes pour capturer leurs dépendances spatiales et structurelles.

Les transformeurs (*transformers*), architecture émergente particulièrement performante ces dernières années, utilisent un mécanisme d'auto-attention qui permet d'estimer l'importance relative de différentes parties

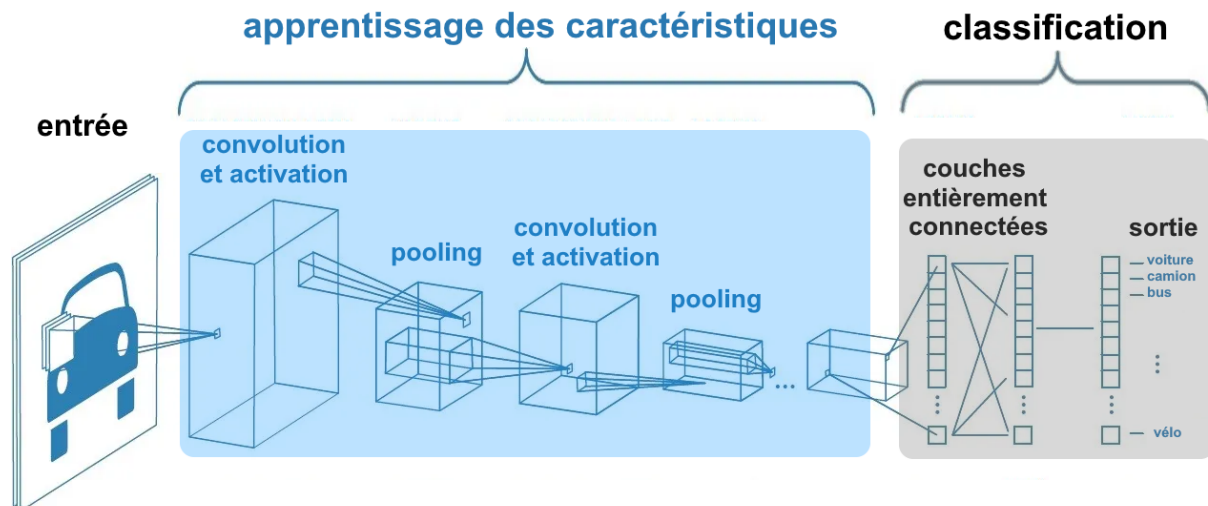
[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[68] : GOODFELLOW et al. (2018), *L'ap-  
prentissage profond / Ian Goodfellow, Yo-  
shua Bengio, Aaron Courville*

[66] : GÉRON (2017), *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow*

[68] : GOODFELLOW et al. (2018), *L'ap-  
prentissage profond / Ian Goodfellow, Yo-  
shua Bengio, Aaron Courville*

[74] : LINDSAY (2021), « Convolutional Neural Networks as a Model of the Visual System »



**FIG. 3.5 : Architecture d'un CNN** En bleu, la partie d'apprentissage des caractéristiques, en gris la partie de classification (*modifié, d'après Sumit Saha*)

[75] : VASWANI et al. (2017), « Attention is All you Need »

des variables d'entrée ce qui leur permet d'apprendre des dépendances à long terme plus efficacement que les RNNs et les LSTMs. Cela les rend particulièrement performants pour des tâches de NLP ou de traduction de texte [75].

Les tâches des modèles présentés ici ont pour point commun de nécessiter une sortie qui réduit l'information la dimensionnalité des données d'entrée. Par exemple, en classification d'image, le but est d'obtenir l'appartenance à une classe parmi une collection de classes. Concrètement, l'entrée est une image correspondant à une matrice de plusieurs milliers de paramètres, tandis que la sortie est un vecteur contenant la probabilité d'appartenance à quelques classes. Cependant, l'apprentissage profond permet aussi de fabriquer de nouvelles données contenant autant d'informations que les données d'apprentissages.

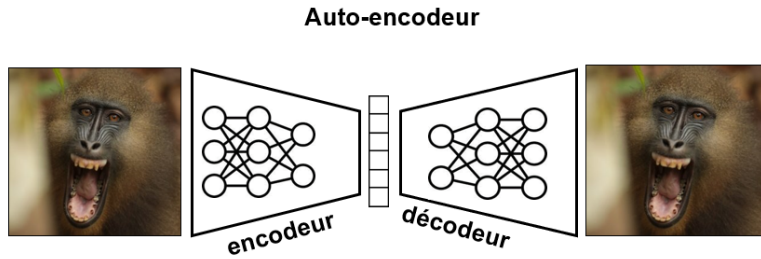
### 3.1.3 Intelligence artificielle générative

[76] : ZHOU et al. (2024), *A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming*

[77] : RUFF et al. (2021), « AlphaFold and Implications for Intrinsically Disordered Proteins »

L'intelligence artificielle générative (IA générative) désigne une classe de modèles d'IA capables de créer de nouvelles données à partir d'un ensemble d'apprentissage. Les modèles génératifs apprennent à comprendre et à reproduire la distribution des données d'entraînement pour générer de nouvelles instances semblables. Ces pratiques peuvent être à usage créatif, en générant des vidéos, sons ou images, à but informatif, en synthétisant ou paraphrasant des données [76], mais aussi dans un but de recherche scientifique. Nous citerons par exemple la génération de nouvelles structures moléculaires pour des composés pharmaceutiques potentiels [77], ou encore la synthèse de stimuli pour des expériences comportementales, que nous développerons en partie 4.2.2 Au-delà de ces cas particuliers, nous discuterons à la fin de ce manuscrit de l'intérêt de l'IA générative en écologie et en évolution, et des possibilités que ces technologies émergentes ouvrent dans ces domaines.

Les auto-encodeurs (AE) sont une des familles d'architectures d'IA générative basées sur de l'apprentissage profond. Il s'agit d'une architecture en

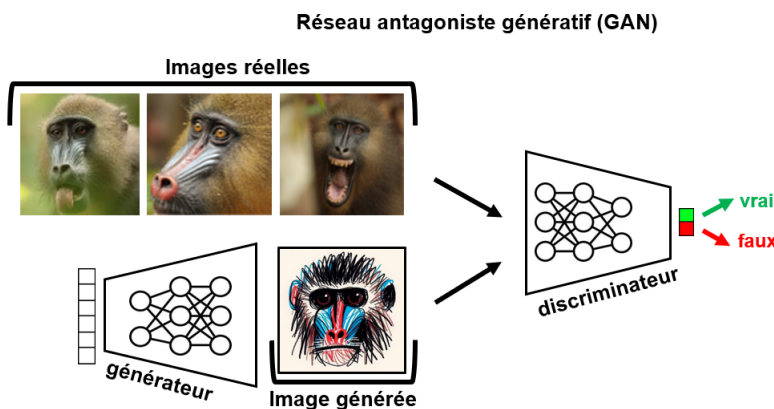


**FIG. 3.6 : Architecture schématique d'un auto-encodeur (production personnelle)**

2 parties : un encodeur réduit la dimensionnalité d'une donnée d'entrée en un vecteur latent, comme le ferait un réseau de neurones classique. Ensuite, à partir de ce vecteur, un décodeur essaie de reconstruire la donnée d'entrée en minimisant la différence entre l'information générée et l'information d'origine (Cf. Fig. 3.6) [78]. Les auto-encodeurs ont initialement été conçus comme une méthode de réduction de la dimensionnalité des données d'entrée, plus que comme une méthode générative, et les données synthétiques créées ne sont pas nécessairement réalistes. Une version plus avancée, les auto-encodeurs variationnels (VAEs, Variational Auto Encoders) fonctionne sur le même principe mais a spécialement pour objectif de générer de nouvelles données. Pour cela, ils encodent les données d'entrée non pas en un seul vecteur mais en une distribution gaussienne latente, décrite par une moyenne et une variance. Cela leur permet d'être bien plus performants pour générer des données synthétiques réalistes [79].

[78] : KRAMER (1991), « Nonlinear principal component analysis using autoassociative neural networks »

[79] : KINGMA et al. (2022), *Auto-Encoding Variational Bayes*



**FIG. 3.7 : Architecture schématique d'un réseau antagoniste génératif (GAN, Generative Adversarial Network) (production personnelle)**

Les réseaux antagonistes génératifs (GANs, Generative Adversarial Networks)[80] sont aussi une architecture en deux parties, mais articulées sur un fonctionnement différent de celui des auto-encodeurs. Le générateur essaie de fabriquer des données synthétiques qui imitent les données d'apprentissage réelles. Le discriminateur prend des données soit réelles soit fabriquées par le générateur comme entrée et essaie de distinguer entre les deux (Cf. Fig. 3.7). L'avantage des GANs est qu'ils n'ont pas besoin de paires étiquetées de données d'entrée et de sortie, ce qui permet d'utiliser de grandes quantités de données non étiquetées. Grâce à ces mécanismes, ils sont capables de produire des données réalistes synthétiques. Apparus en 2014, les GANs ont depuis connu des améliorations majeures : en 2017, leur réalisme a été amélioré [81] à partir de travaux d'optimisation basé sur le transport optimal [82]. En 2019, une version plus complexe, ajoutant une troisième partie liée au générateur, StyleGAN, a permis d'améliorer

[80] : GOODFELLOW et al. (2014), *Generative Adversarial Networks*

[81] : ARJOVSKY et al. (2017), *Wasserstein GAN*

[82] : OLLIVIER et al. (2014), *Optimal Transport*

[83] : KARRAS et al. (2019), *A Style-Based Generator Architecture for Generative Adversarial Networks*

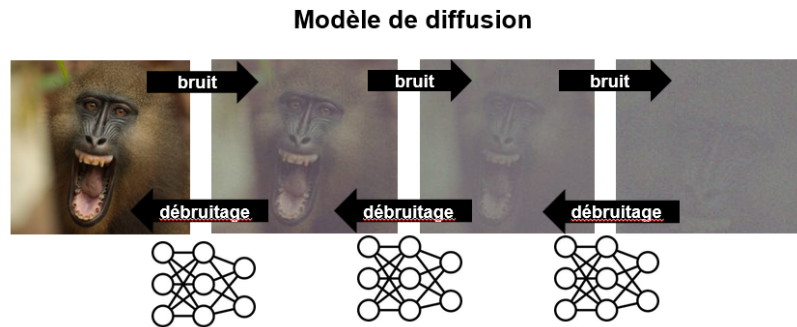
[84] : BERMANO et al. (2022), « State-of-the-Art in the Architecture, Methods and Applications of StyleGAN »

[85] : WU et al. (2020), *StyleSpace Analysis*

[86] : KARRAS et al. (2021), *Alias-Free Generative Adversarial Networks*

fortement le réalisme des données synthétiques, en particulier concernant les visages [83]. Cette troisième partie, appelée réseau de cartographie (*mapping network*), permet de créer des vecteurs latents dont les variations linéaires sont plus alignées avec les variations sémantiques dans les images générées [84]. En conséquence, il devient plus facile de manipuler des caractéristiques spécifiques de l'image, comme l'identité, la pose ou l'éclairage [85]. C'est cette architecture qui est derrière le site "This person does not exist". Des versions encore plus performantes de StyleGAN sont apparues : StyleGAN2 en 2020 puis StyleGAN3 en 2021 [86].

FIG. 3.8 : Architecture schématique d'un modèle de diffusion (production personnelle)



[87] : SOHL-DICKSTEIN et al. (2015), *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*

Les diffuseurs, ou modèles de diffusion (*diffusion models*) sont une architecture apparue en 2015 [87]. Ces modèles fonctionnent en deux étapes. Tout d'abord, chaque donnée d'entrée est subdivisée en plusieurs instances de plus en plus bruitées, jusqu'à du bruit pur : c'est le processus de diffusion. Ensuite, un réseau apprend à éliminer progressivement le bruit pour reconstruire les données à partir du bruit (Cf. Fig. 3.8). Le modèle final apprend donc à générer des données réalistes à partir de bruit.

Les transformeurs (*transformers*) précédemment évoqués peuvent aussi être utilisés dans des architectures génératives, comme GPT, l'un des modèles derrière ChatGPT par exemple.

En discussion générale, nous développerons l'intérêt des techniques d'IA générative en écologie et en évolution, et des perspectives ouvertes par ces avancées.

### 3.2 Un modèle et des métriques

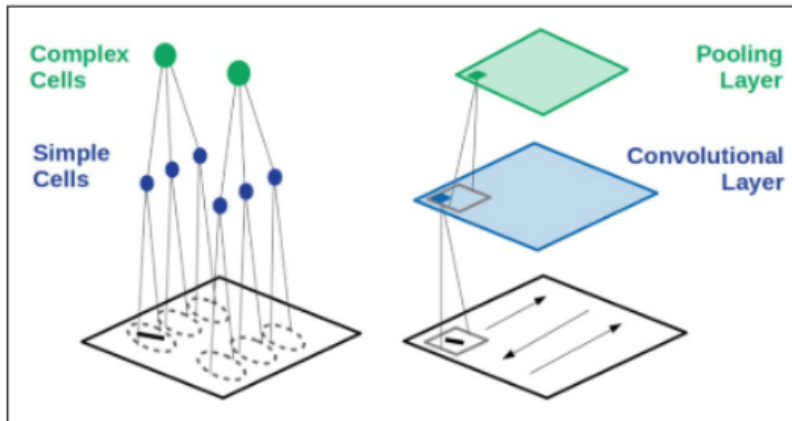
On a vu en partie 2.2.2 que la beauté pouvait être expliquée par la fluence, une sensation de facilité dans le traitement de l'information par le système visuel qui provoque un plaisir. Pour comprendre et valider ce mécanisme, et au-delà, l'attractivité des visages, cela implique deux choses : un modèle du système visuel et une métrique pour mesurer l'attractivité, la beauté ou a fortiori la fluence.

Les premiers travaux sur la modélisation de la fluence l'ont approchée par des mesures de caractéristiques de stimuli visuels, comme la symétrie ou le contraste, considérées comme des proxy plausibles [88]. Toutefois, ces approches sont incomplètes car elles ne prennent en compte que les caractéristiques des stimuli. Nous avons en effet vu en 1.2.2 que la fluence était liée fondamentalement à l'interaction entre le stimuli et l'observateur.

[88] : MAYER et al. (2018), « Quantifying Visual Aesthetics Based on Processing Fluency Theory »

Des mesures plus pertinentes doivent donc concerner la manière dont les informations des stimuli sont traitées dans le système visuel.

Pour modéliser le système visuel, il y a deux approches. L'approche fonctionnelle cherche à correspondre uniquement aux entrées et aux sorties du système modélisé. A l'inverse, l'approche mécanistique cherche à modéliser en plus les composants du système [89]. Dans le système



[74] : LINDSAY (2021), « Convolutional Neural Networks as a Model of the Visual System »

**FIG. 3.9 : Relation entre les composants du système visuel et les opérations de base d'un réseau de neurones convolutif (CNN)** Les cellules simples (en bleu, à gauche) du système visuel sont homologues aux couches de convolution des CNN (en bleu, à droite), et les cellules complexes (en vert, à gauche) du sont homologues aux couches de pooling des CNN (en vert, à droite) (d'après [74])

visuel décrit en 1.1.1, à un niveau de description plus précis que les aires que nous avons présentées, les cellules simples sont spécialisées dans la détection de caractéristiques spécifiques, et les cellules complexes, en plus petit nombre, agrègent les information issues de plusieurs cellules simples (Fig. 3.9). Ce processus est répété plusieurs fois en série, les sorties de cellules complexes étant les entrées des cellules simples de l'aire suivante. Ce processus a fait l'objet de modèles comme HMAX [90], mais plus récemment, il a été montré que les CNN, décrits en partie 3.1.2, étaient des modèles particulièrement performants du système visuel [74, 91]. En effet, ils ont des points communs structurels : le diptyque détection des caractéristiques - agrégation de l'information correspond à la fois aux cellules simples puis complexes des aires du système visuel, et aux couches de convolutions et de pooling des CNN (Cf. Fig. 3.9). Les différentes couches des CNN peuvent alors être comparées aux aires du système visuel, ou le flux d'information gagne en complexité et en abstraction à mesure qu'il passe à travers le réseau. Pour valider ce modèle, des travaux ont montré qu'il y avait une corrélation entre les activations des CNN et l'activité neuronale du cortex visuel [92]. Plus précisément, les activations de la dernière couche des CNN corrélaient avec l'activité du cortex inféro-temporal, et l'avant dernière couche prédit l'aire V4. Les architectures de CNN avec lesquelles cette analogie fonctionne le plus sont le plus souvent des architectures peu profondes, avec moins d'une vingtaine de couches comme AlexNet [93], qui a 8 couches ou VGG16 [94], qui en a 16. Cela fonctionne moins bien dans la plupart des cas avec les architectures plus profondes même si dans certaines situations, elles modélisent bien les boucles de rétroaction qui peuvent exister dans le système visuel [95].

A partir d'un tel modèle, des métriques appropriées doivent aussi être utilisées pour modéliser la fluence. Nous avons vu en partie 2.3.1 que dans le cadre du choix de partenaire, une préférence pouvait être sélectionnée pour répondre à un signal donné, mais aussi que, dans certains cas, l'apparition de la préférence pouvait pré-dater celle du signal. De telles préférences restent donc latentes jusqu'à ce que le signal apparaisse,

[89] : KAY (2018), « Principles for models of neural information processing »

[90] : RIESENHUBER et al. (1999), « Hierarchical models of object recognition in cortex »

[74] : LINDSAY (2021), « Convolutional Neural Networks as a Model of the Visual System »

[91] : KRIEGESKORTE (2015), « Deep neural networks »

[92] : YAMINS et al. (2014), « Performance-optimized hierarchical models predict neural responses in higher visual cortex »

[93] : KRIZHEVSKY et al. (2012), « ImageNet Classification with Deep Convolutional Neural Networks »

[94] : SIMONYAN et al. (2015), *Very Deep Convolutional Networks for Large-Scale Image Recognition*



[95] : KAR et al. (2019), « Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior »

[96] : RYAN et al. (1993), « Sexual Selection and Signal Evolution »

[97] : RENOULT et al. (2019), « Processing bias »

[98] : SIMONCELLI et al. (2001), « Natural image statistics and neural representation »

[99] : SPEHAR et al. (2015), « Beauty and the beholder »

[100] : SPEHAR et al. (2016), « Taxonomy of Individual Variations in Aesthetic Responses to Fractal Patterns »

[101] : REDIES et al. (2007), « Artists portray human faces with the Fourier statistics of complex natural scenes »

[101] : REDIES et al. (2007), « Artists portray human faces with the Fourier statistics of complex natural scenes »

[102] : WINKIELMAN et al. (2012), « Fluency of consistency »

[97] : RENOULT et al. (2019), « Processing bias »

[103] : HOLZLEITNER et al. (2019), « Comparing theory-driven and data-driven attractiveness models using images of real women's faces »

[104] : WINKIELMAN et al. (2006), « Prototypes Are Attractive Because They Are Easy on the Mind »

[105] : REBER et al. (1998), « Effects of Perceptual Fluency on Affective Judgments »

éventuellement. Les biais perceptuels peuvent expliquer ces préférences sexuelles latentes, en tant que sous-produits de l'adaptation des systèmes perceptifs à des tâches non liées à la sélection sexuelle [96]. Le biais de codage efficient (*efficient coding*) [97] est une hypothèse de mécanisme neuronal de la fluence. L'efficacité est ici définie comme le traitement de l'information avec un usage économique des ressources. Autrement dit, la fluence est un biais d'attractivité qui intervient lorsqu'il y a un codage efficient de l'information dans le système visuel, provoquant une sensation de facilité car le système visuel est adapté pour ce traitement économique de l'information.

Ce biais explique beaucoup de résultats empiriques sur l'attractivité, et notamment des préférences pour des caractéristiques visuelles identiques à celles que l'on retrouve dans la nature, donc auxquelles est adapté le cerveau [98]. Un exemple bien documenté est celui des préférences visuelles pour l'invariance à l'échelle étudiée à travers la pente de Fourier. La pente de Fourier est une manière de décrire comment les détails d'une image se décomposent en différentes fréquences spatiales. D'un part, des travaux ont montré que le système visuel est le plus sensible au degré d'invariance à l'échelle qui est le plus fréquent dans la nature [99]. D'autre part, d'autres travaux ont montré que des stimuli visuels étaient jugés d'autant plus attractifs que leur degré d'invariance à l'échelle est proche de celui observé le plus fréquemment dans la nature [100]. Cette attraction pour une invariance à l'échelle "naturelle" a été proposée pour expliquer, par exemple, que les portraits artistiques (de différentes époques, styles, origines culturelles) ont tendance à avoir le même degré d'invariance à l'échelle que celui observé le plus fréquemment dans des paysages naturels, qui diffère du degré d'invariance à l'échelle observé le plus fréquemment dans des portraits photographiques non artistiques (e.g., photographies d'identité [101]).

Pour mesurer l'efficacité du traitement de l'information dans un CNN, il est nécessaire de la caractériser de manière plus opérationnelle. En mathématiques, la sparsité est la situation où la majorité des éléments d'une structure, telle qu'un vecteur ou une matrice, sont égaux à zéro ou ont une valeur très faible, tandis que peu d'éléments ont des valeurs très élevées. En terme de perception visuelle, un stimuli sparse active peu de neurones simultanément, permettant donc un traitement efficient de l'information. Ainsi, la sparsité peut être un moyen de modéliser la fluence [101, 102]. Elle a d'ailleurs été utilisée pour prédire l'attractivité des visages [97], et est même un meilleur prédicteur que le dimorphisme sexuel, la centralité et la corpulence [103].

On a vu en partie 2.2.2 que les prototypes, représentants moyens d'une catégorie, sont préférés par rapport aux stimuli moins centraux. La fluence peut expliquer cette préférence pour les prototypes [104]. En effet, le système visuel a évolué pour les traiter à moindre coup étant donné leur familiarité, ce qui permet de les catégoriser rapidement et précisément. Plus généralement, le concept de typicalité décrit dans quelle mesure un stimulus correspond à une représentation centrale, autrement dit à quel point un stimulus est un prototype, et ses liens avec la fluence sont établis [105]. Cette typicalité peut être calculée comme la vraisemblance (*likelihood*) d'une image par rapport à une distribution de référence. Des travaux ont montré que la typicalité de visages, ainsi calculée, pouvait en partie prédire leur attractivité [106].

La typicalité peut aussi influencer la perception du dimorphisme sexuel d'une espèce, que l'on a évoqué en partie 2.3.2. Plus la différence entre les mâles et les femelles d'une espèce est importante, plus les prototypes qui caractérisent les représentations mentales de chacun des deux sexes seraient alors différents. Pour un même sexe, il y a deux manières de quantifier cela, au regard à la fois de la typicalité et du dimorphisme. Par exemple, quand on étudie la féminité, on peut l'évaluer par la centralité, c'est à dire à quel point une femelle est représentative de la classe femelle, ou par la *femaleness*<sup>3</sup>, à quel point une femelle est différente de la classe des mâles.

[106] : BRIELMANN et al. (2022), « A computational model of aesthetic value »



# L'intelligence artificielle pour comprendre expérimentalement l'attractivité faciale

# 4

Pour comprendre l'attractivité, il est certes possible de la modéliser à travers l'une de ses composantes, la beauté, comme nous l'avons vu précédemment. Néanmoins, une autre manière de la comprendre est d'observer et de mesurer empiriquement, les comportements et les interactions socio-sexuelles de vrais animaux. Cette attractivité est un facteur clé du choix de partenaire décrit en partie 2.3.

Pour ce faire, il est possible d'enregistrer les choix dans la nature et d'étudier, par une approche corrélative, quels phénotypes semblent attractifs. Cela a l'avantage de pouvoir mettre en relation la complexité des choix de partenaires, avec l'environnement social et écologique des individus. Par exemple, des travaux ont montré que, chez plusieurs espèces de mammifères, les femelles préfèrent les mâles dominés mais les interactions avec eux sont limitées par un comportement coercitif des mâles dominants [107]. Cette approche permet également de formuler des hypothèses sur le choix de partenaire à partir d'observations comportementales.

Pour tester ces hypothèses, outre les analyses corrélatives il est possible d'avoir recours à des expériences comportementales. Dans cette section, nous allons d'abord expliquer en quoi consiste l'approche expérimentale et quelles sont ses particularités dans le cadre de l'étude de signaux visuels en 4.1. Ensuite, en partie 4.2, nous discuterons de ce que peut apporter l'intelligence artificielle générative pour des expériences étudiant les signaux visuels.

## 4.1 L'approche expérimentale

### 4.1.1 Généralités

L'approche expérimentale consiste à concevoir et à mener des expériences en conditions contrôlées pour examiner les relations de cause à effet entre différentes variables. Pour commencer, il faut formuler une hypothèse, une prédiction, basée sur des théories existantes ou des observations préalables. Ensuite, la conception de l'expérience *stricto sensu* implique de planifier comment elle sera menée en définissant tous les paramètres ou les séries de paramètres. Il faut définir les variables de réponse (ce qui est mesuré) et les variables explicatives (ce qui est manipulé ou contrôlé), comme les caractéristiques des stimuli, les individus étudiés ou le temps et le nombre d'essais à réaliser selon ces paramètres. Dans la mesure du possible, il faut standardiser les conditions de l'expérience, ce qui est plus ou moins facile en fonction du contexte, en particulier quand les expériences concernent des animaux. Pendant le déroulement de l'expérience, les données expérimentales sont recueillies, soit par des capteurs, soit par un observateur. Ces données sont ensuite traitées, puis analysées. L'approche expérimentale comporte plusieurs avantages. Elle permet le contrôle de facteurs de confusion, qui ne concernent pas directement les hypothèses mais qui peuvent influencer à la fois la variable de réponse et les variables

4.1 L'approche expérimentale . . . . .	27
4.1.1 Généralités . . . . .	27
4.1.2 Ecologie sensorielle et écologie visuelle . . . . .	28
4.1.3 Construire un protocole pour tester une préférence visuelle . . . . .	28
4.2 L'intelligence artificielle générative comme outil expérimental . . . . .	30
4.2.1 Fabriquer des stimuli . . . . .	31
4.2.2 Apport de l'intelligence artificielle générative . . . . .	32

[107] : WONG et al. (2005), « How is female mate choice affected by male competition ? »

explicatives, rendant difficile l'interprétation de résultats d'observation comportementales réalisées en dehors d'un cadre expérimental. Ainsi, elle permet de renforcer la robustesse des analyses sur l'existence de liens causaux en plus d'être corrélatifs. De plus, les individus qui existent réellement dans la nature ne représentent qu'un échantillon de toutes les combinaisons de phénotypes possibles. Manipuler des stimuli permet de mettre en évidence des préférences pour des phénotypes réels mais qui n'auraient pas été détectés dans la nature, ou pour des phénotypes qui n'existent pas (c'est-à-dire, des préférences latentes ; Cf. partie 2.3).

Néanmoins, l'approche expérimentale présente aussi des limites. Les conditions contrôlées des expériences peuvent ne pas refléter les situations réelles, et donc ne pas être généralisables à des contextes réels. Il est aussi difficile d'identifier les variables qui doivent être contrôlées et cela peut compliquer l'interprétation des résultats. Les participants, humains ou animaux, peuvent représenter, justement du fait qu'ils ont souhaité ou été sélectionnés pour y participer, un échantillon qui n'est pas représentatif de la population que l'on veut étudier. De plus, leurs réactions en condition expérimentale peuvent être influencées par le fait qu'ils ont conscience de l'expérience, ce qui peut biaiser les résultats. Enfin, mesurer des variables comportementales de manière précise peut être difficile et les instruments de mesure ou l'analyse par un observateur peuvent introduire des erreurs ou des biais.

Quoi qu'il en soit, un protocole expérimental doit être conçu en fonction des spécificités des champs disciplinaires des hypothèses qu'il veut tester. Dans le cadre de la perception de visages, nous nous plaçons dans un contexte d'écologie visuelle.

### 4.1.2 Écologie sensorielle et écologie visuelle

La perception qu'ont les espèces animales de leur environnement est modulée par l'expérience sensorielle qu'elles en ont. L'écologie sensorielle est la discipline qui étudie comment les êtres vivants acquièrent, traitent et utilisent cette information sensorielle, et comment cela influence des changements évolutifs [108].

L'écologie visuelle est la branche de l'écologie sensorielle qui étudie spécifiquement les signaux visuels, dont nous avons détaillé la perception en partie 2.1.1. Un visage peut être vu comme un ensemble de caractéristiques visuels. Celles-ci peuvent être des couleurs, des motifs, des textures, mais aussi des mouvements [109] et peuvent avoir un rôle de signal, qui a évolué.

### 4.1.3 Construire un protocole pour tester une préférence visuelle

Pour étudier l'attractivité dans un cadre expérimental, un type de protocole adapté est le test de choix. Dans un tel test, les participants doivent choisir parmi plusieurs options présentées simultanément ou successivement. Ces tests peuvent varier en complexité, allant de choix simples entre deux options à des choix complexes impliquant un plus grand nombre d'alternatives [110]. Cela permet d'analyser les préférences et les processus

[108] : STEVENS (2013), *Sensory Ecology, Behaviour, and Evolution*

[109] : FRANK (2015), « Visual Ecology Thomas W. Cronin, Sönke Johnsen, N. Justin Marshall, and Eric J. Warrant, editors »

[110] : DELAITRE et al. (2023), « Female great tits (*Parus major*) reproduce earlier when paired with a male they prefer »

de prise de décision en temps réel.

Il faut différencier les notions de choix et de préférence. Une préférence est une inclination pour une option par rapport à une autre basée sur des désirs ou des besoins. Un choix est l'acte de sélectionner une option parmi plusieurs alternatives disponibles, mais ce concept ne présume pas des raisons qui ont poussé à l'acte. C'est de fait essentiel de considérer ces différences dans l'interprétation d'un test de choix. Si il y a une différence de choix entre deux types de stimuli, cela veut dire que les individus sont capables de discriminer les objets selon la variable qui décrit la différence entre les deux [111]. Un tel résultat doit donc être mis en perspective avec une interprétation et un contexte biologique ou comportemental pour présumer plus que cette capacité de discrimination. Ainsi, la question de déterminer si des signaux provoquent une attirance est intrinsèquement liée avec l'interprétation des comportements qui expriment cette préférence.

Le paradigme du temps d'observation est une manière d'interpréter des choix issus de tests concernant des stimuli visuels. Une tâche de temps d'observation consiste à présenter plusieurs stimuli visuels à des sujets et à mesurer leur temps d'observation de chacun des stimuli [112]. Il y a trois types de tâches. Une première tâche est l'habituation visuelle, qui consiste à montrer un stimulus de manière répétée jusqu'à ce qu'il y ait une diminution de la réaction, puis à montrer un nouveau stimulus pour évaluer s'il y a une augmentation de la réaction. Une seconde tâche est la violation de l'attente, qui consiste à montrer au sujet des stimuli qui sont ou non conforme à des attentes présumées. Si un stimulus supposé inattendu est regardé plus longtemps, cela confirme que ce stimulus est effectivement inattendu. Enfin, une troisième tâche est la tâche de biais visuel, qui consiste à montrer plusieurs stimuli simultanément ou séquentiellement et regarder celui qui est préféré. Montrer les stimuli simultanément permet de comparer plus rigoureusement les réactions, mais l'approche séquentielle est utile quand il y a une combinaison de modalités décrivant les stimuli à choisir [113].

On peut mesurer le temps d'observation de plusieurs manières. Un observateur peut le compter visuellement pendant l'essai, ce qui est simple à mettre en place, mais risque d'être imprécis et peut biaiser l'essai si l'observateur est visible. Un capteur vidéo peut enregistrer les essais pour qu'ils soient encodés a posteriori. Cela permet d'améliorer la précision des annotations. Enfin, des capteurs d'*eye tracking* peuvent enregistrer finement où se dirige le regard des individus. Cela a l'avantage d'être automatique et précis, mais est plus facile à mettre en place chez des humains que chez des animaux, principalement pour des questions d'étalonnage [113]. De fait, le paradigme du temps d'observation a historiquement d'abord été utilisé pour des expérimentations avec des enfants humains trop jeunes pour verbaliser leurs préférences, mais est de plus en plus utilisé chez les animaux, en particulier chez les primates [113]. Ainsi, les jeunes primates présentent plus de curiosité pour les stimuli visuels dans ce genre d'expériences car ils les regardent plus longtemps que des primates adultes [111]. Les stimuli peuvent être de plusieurs types : il peut s'agir d'objets concrets, par exemple des oeufs, de représentations fidèles de ces objets concrets, par exemple des photos d'oeufs, ou de représentations synthétiques de ces objets, par exemple des maillages tridimensionnels représentant des oeufs. Des travaux ont montré que les animaux réagissent moins à des images

[111] : PFEFFERLE et al. (2014), « Monkeys Spontaneously Discriminate Their Unfamiliar Paternal Kin under Natural Conditions Using Facial Cues »

[112] : SPELKE (1985), « Preferential-looking methods as tools for the study of cognition in infancy. »

[113] : WINTERS et al. (2015), « Perspectives »

[113] : WINTERS et al. (2015), « Perspectives »

[113] : WINTERS et al. (2015), « Perspectives »

[111] : PFEFFERLE et al. (2014), « Monkeys Spontaneously Discriminate Their Unfamiliar Paternal Kin under Natural Conditions Using Facial Cues »

[114] : WAITT et al. (2003), «Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice»

[113] : WINTERS et al. (2015), «Perspectives»

[115] : LEWIS et al. (2023), «Bonobos and chimpanzees remember familiar conspecifics for decades»

d'objets que lorsqu'ils sont directement exposés aux objets directement. Dans le cas des primates, c'est encore plus marqué quand les objets sont d'autres primates [114]. Comme tout test de choix, les tests réalisés sous le paradigme du temps d'observation sont insuffisants pour comprendre les raisons biologiques expliquant les résultats. Ils doivent être mis au regard d'autres mesures pendant les expérimentations, par exemple une annotation de comportements particuliers, et d'autres approches au-delà de l'approche expérimentale, comme des études corrélatives réalisées sur des observations hors contexte expérimental. En effet, on ne peut donc pas parler de préférence pour un stimulus dans ce contexte mais simplement de discrimination entre les stimuli : le terme de biais visuel est alors à privilégier pour parler du stimulus le plus regardé [113]. Ce paradigme peut à la fois être mis en place en conditions contrôlées dans des pièces fermées avec des animaux en captivité (Cf. Fig. 4.1), mais aussi en terrain ouvert avec des animaux sauvages (Cf. Fig. 4.2).

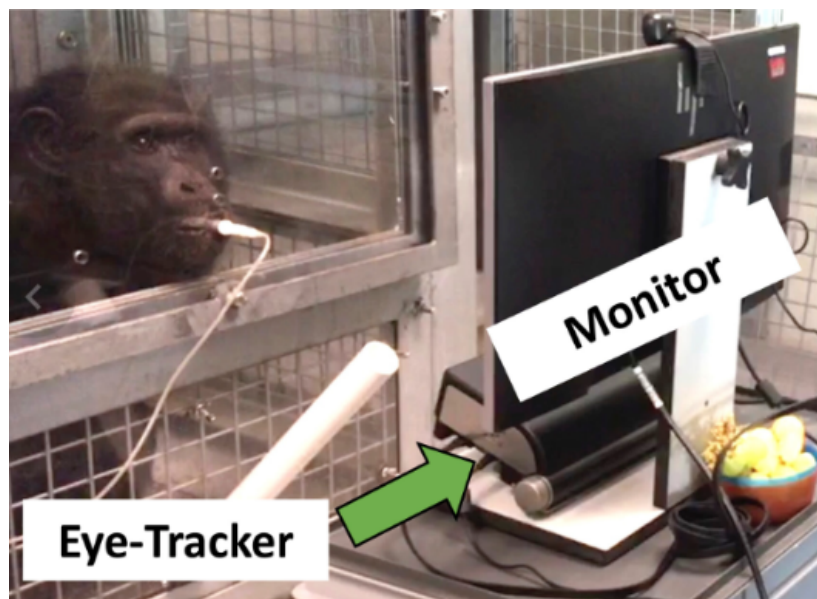


FIG. 4.1 : Un test de choix réalisé avec des animaux en captivité et un dispositif contrôlé (d'après [115])

[116] : ROSENFELD et al. (2019), «Experimental evidence that female rhesus macaques (*Macaca mulatta*) perceive variation in male facial masculinity»

Des tests de choix basés sur le paradigme du temps d'observation peuvent apparaître comme pertinents dans l'évaluation de préférences de typicalité féminine ou masculine, évoquée en partie 2.2.2. Des travaux s'intéressant aux préférences pour les individus apparentés ont montré que chez certains primates, ce choix était d'autant plus important quand les individus choisis étaient du même sexe que les individus testés [111]. De plus, chez des macaques rhésus femelles, des tests de choix ont été réalisés sur des photographies d'individus mâles plus ou moins masculins. Ces travaux ont montré un biais visuel plus important en faveur des mâles les plus masculins, confirmant que ce type de facteur peut être discriminant.

## 4.2 L'intelligence artificielle générative comme outil expérimental

Nous avons vu qu'en approche expérimentale, il est possible d'utiliser des stimuli visuels pour voir comment des sujets réagissent. Pour faire cela, des stimuli qui existent déjà peuvent être utilisés, ou des représentations de ces stimuli (comme des photos). Il faut alors créer un protocole

expérimental prenant en compte les caractéristiques de ces stimuli pour tester des hypothèses [117, 118]. Cela est limité car la combinaison des



FIG. 4.2 : Un test de choix réalisé avec des animaux sauvages (d'après [116])

caractéristiques de ces stimuli est finie. Dans la nature, l'ensemble des combinaisons des phénotypes possibles est beaucoup plus grand. De plus, la variation entre des stimuli différents peut être composite et ne pas concerner uniquement les variables permettant de tester les hypothèses, ce qui induit des facteurs de confusion.

Pour pallier ce problème, il est possible de fabriquer des stimuli de synthèse en manipulant précisément des caractéristiques d'intérêt. Pour ce qui est des stimuli visuels, comme le sont les visages, nous verrons en partie 4.2.1 qu'on peut faire cela avec des méthodes traditionnelles de vision par ordinateur. Ces méthodes permettent d'éviter certains des écueils mentionnés, mais nous expliquerons en partie 4.2.2 qu'utiliser l'IA générative permet d'aller encore plus loin.

### 4.2.1 Fabriquer des stimuli

C'est principalement dans le domaine des sciences comportementales humaines que des expériences basées sur des stimuli synthétiques ont été réalisées. Par exemple, dans le domaine de la reconnaissance des apparentés, des travaux [119] ont créé des visages proches de visages réels pour simuler une variation dans le degré d'apparentement. En effet, obtenir une base de données de visages de personnes réelles apparentées est difficile. Des scores d'apparentement ont été calculés selon un panel d'observateur, et il n'y avait pas de différence de score concernant des apparentés réels ou des apparentés synthétiques, ce qui montre la pertinence de cette approche. De même, une étude a manipulé des portraits d'hommes pour modifier leur masculinité afin de tester des hypothèses de psychologie évolutionniste [120] (Cf. Fig. 4.3).

Dans ce domaine, c'est très majoritairement le logiciel *Psychomorph* [121] spécialisé dans les visages qui est utilisé pour créer des stimuli. La méthode consiste à modifier des stimuli déjà existants à partir d'un ensemble de repères faciaux (*landmarks*) qui auront été placés manuellement ou

[111] : PFEFFERLE et al. (2014), « Monkeys Spontaneously Discriminate Their Unfamiliar Paternal Kin under Natural Conditions Using Facial Cues »

[117] : FIALA et al. (2021), « Facial attractiveness and preference of sexual dimorphism »

[118] : WINTERS et al. (2019), *The structure of species discrimination signals across a primate radiation*

[119] : BOUSQUET et al. (2022), « Transforming faces to mimic natural kin »

[120] : NILA et al. (2019), « Male Homosexual Preference »



automatiquement, en faisant varier ces repères. Il s'agit de *morphing*. L'ensemble des variables décrivant ces repères peut être vu géométriquement comme un espace multidimensionnel où chaque dimension est un repère, et chaque point un visage placé dans l'espace en fonction des valeurs de tous ses repères. Ce type d'espace est appelé espace des visages (*face space*). Des méthodes appelées procrustes permettent a posteriori d'améliorer les stimuli en éliminant les différences entre les visages liés à la position, à l'échelle et à la rotation.

Chez d'autres espèces que les humains, créer des stimuli synthétiques n'est pas très répandu. Néanmoins, des approches alternatives existent pour compenser les limites de l'utilisation de stimuli réels. Des travaux chez les primates [118] ont identifié les pixels de fortes variations de variables d'intérêt sur des portraits réels de cercopithèques permettant ainsi de sélectionner au mieux les stimuli adaptés à une expérience permettant de tester l'impact de ces variables. Néanmoins, les méthodes utilisées étaient basées sur des combinaisons linéaires des pixels des images ne prenant pas en compte les variations inhérentes à ce type de stimuli.

[120] : NILA et al. (2019), « Male Homosexual Preference »

#### 4.2.2 Apport de l'intelligence artificielle générative

L'IA générative permet d'aller plus loin dans la création de stimuli. Tout d'abord, nous avons vu en partie 3.1 que ces techniques permettent de capter des relations complexes entre des variables, et de déterminer leur distribution. Cela permet de créer des données suffisamment réalistes pour qu'elles soient difficilement discernables de données réelles, et ce particulièrement pour des images.

Dans le contexte de la création de stimuli, des besoins supplémentaires existent. Les GANs en particulier ont montré leur efficacité pour créer des données réalistes. Ces données sont projetées dans un espace latent de très grande dimension, créé par le réseau. Néanmoins, contrairement aux auto-encoders, pourtant moins performants, les GANs ne permettent en aucun cas de projeter des données réelles dans cet espace latent, qui ne sert qu'à la construction de données synthétiques. Dans un contexte expérimental, il est pourtant primordial de pouvoir comparer des données synthétiques à des données réelles (ou à une distribution de certaines caractéristiques de données réelles) pour pouvoir y associer plus facilement des métadonnées proches (par exemple, pour une image d'un individu, son âge et son sexe).

[121] : ROWLAND et al. (1995), « Manipulating Facial Appearance Through Shape and Color »

##### Encadré 4 : StyleGAN : spécificités de l'architecture

L'architecture StyleGAN permet d'obtenir des images synthétiques particulièrement réalistes et de pallier à certaines limites d'architectures de GANs plus traditionnelles. Le générateur de cette architecture est composé de deux réseaux de neurones. Un premier réseau, appelé **réseau de synthèse**, fabrique une image à partir de bruit et non pas d'un vecteur latent comme les GANs traditionnels. Un second, appelé **réseau de cartographie** (*mapping network*), composé de couches de neurones entièrement connectés les un aux autres (*fully connected*), fabrique un vecteur latent  $w$  à partir d'un vecteur latent  $z$ . Ce vecteur latent  $w$  est intégré de manière identique à chaque couche

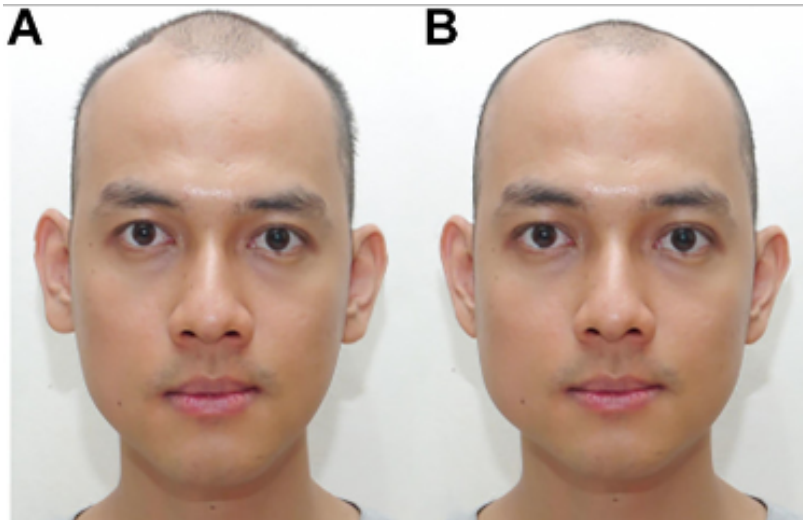


FIG. 4.3 : Un exemple d'image du visage d'un homme rendu plus masculin (A) et plus féminin (B), avec une technique basée sur le morphing (d'après [120] )

du réseau de synthèse (Cf. Fig. 4.4). Ce mécanisme permet de capter plus finement les caractéristiques des images de la base de données d'apprentissage qu'avec un GAN traditionnel : en effet, la distribution de ces caractéristiques dans l'espace  $W$  de l'ensemble des vecteurs  $w$  de toutes les images correspond bien mieux à la distribution réelles que celle de l'espace  $Z$  de l'ensemble des vecteurs  $z$  (Cf. Fig. ??) [83]. Avec un GAN traditionnel, on aurait eu seulement un seul vecteur latent par image, qui aurait eu les mêmes défauts que le vecteur  $z$ .

#### Encadré 5 : StyleGAN : encodage

Les GANs en général et StyleGAN en particulier permettent de générer des images synthétiques associées à un vecteur latent (ou plusieurs dans le cas de StyleGAN) mais pas de connaître le vecteur latent d'une image réelle. Pour cela, il faut ajouter un **encodeur**. Pour StyleGAN, l'encodeur pSp a été proposé [122]. Il s'appuie sur l'idée que l'espace  $W$  n'est pas assez complexe pour bien capter les caractéristiques d'une image extérieure à la base de données d'apprentissage du GAN. De fait, alors que dans le générateur de StyleGAN, le même vecteur  $w$  est injecté dans chaque couche du réseau de synthèse, pSp va, à partir d'une image réelle et suite à un entraînement adéquat, injecter un vecteur latent différent dans chaque couche. La concaténation de tous ces vecteurs  $w$  différents est un vecteur appelé  $w_+$ , que l'on peut positionner dans un espace appelé  $W_+$ . Pour comparer géométriquement des images réelles et des images synthétiques dans  $W_+$ , on peut alors considérer que le vecteur  $w_+$  d'une image synthétique est la concaténation de plusieurs fois son vecteur  $w$ , autant de fois qu'il y a de couches (Cf. Fig. ??).

De plus, des travaux ont montré qu'on pouvait, à travers l'espace latent des GANs, modifier des variables d'intérêt que cet espace parvient à décrire avec succès, et, encore mieux, de manière linéaire [84, 85]. Ainsi, d'après ces conditions, réaliser un morphing tel que décrit dans la partie précédente ne serait possible qu'avec des données synthétiques. Pour solutionner ce problème, des travaux ont proposé des architectures d'encodeurs permettant de projeter des images réelles dans un espace latent

[83] : KARRAS et al. (2019), *A Style-Based Generator Architecture for Generative Adversarial Networks*

[118] : WINTERS et al. (2019), *The structure of species discrimination signals across a primate radiation*

[122] : RICHARDSON et al. (2021), *Encoding in Style*

[83] : KARRAS et al. (2019), *A Style-Based Generator Architecture for Generative Adversarial Networks*

de GAN (voir encadré 5), créant ainsi une architecture encodeur-décodeur entraînée comme un GAN via le mécanisme adversarial, mais s'utilisant plutôt comme un auto-encodeur [84, 122]. L'architecture StyleGAN

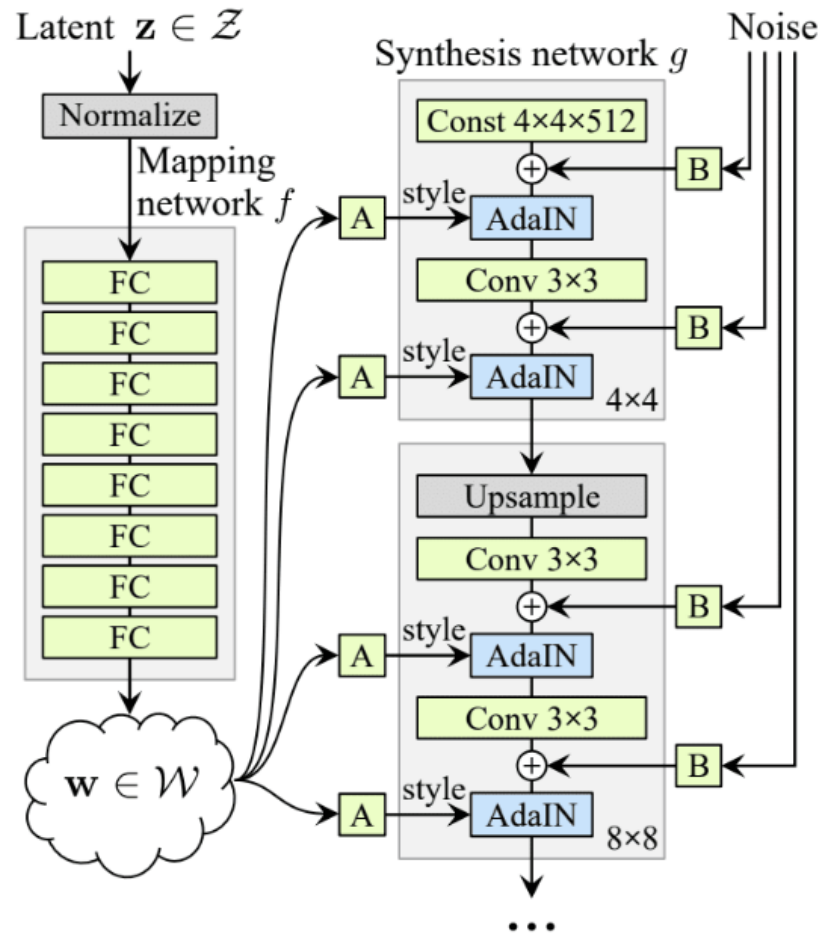


FIG. 4.4 : Architecture du générateur de StyleGAN A gauche, le réseau de cartographie, qui injecte le vecteur  $w$  à chaque couche du réseau de synthèse, à droite. (d'après [83])

répond particulièrement bien à ces enjeux (Voir encadré 4).

Ainsi, l'IA générative peut permettre de créer des stimuli pour des expériences comportementales. Mieux, elle peut permettre d'identifier des variables d'intérêt de manière automatique pour modifier ces stimuli dans la direction qu'elles décrivent, tout en contrôlant la modification d'autres variables qui doivent rester identiques pour éviter de créer des facteurs de confusion.

# Des modèles pour étudier l'attractivité faciale à l'aune de l'intelligence artificielle

## 5

[123] : SOLARI et al. (2007), « Mammal Species of the World »

### 5.1 Force et complémentarités des modèles humains et mandrills

Les mandrills, *Mandrillus sphinx*, et les humains, *Homo sapiens*, sont deux espèces de primates de l'infra-ordre des simiiformes, plus communément appelés singes. Le dernier clade commun les réunissant est le micro-ordre des catarrhiniens, principalement défini par leurs narines dirigées vers le bas [123].

Nous choisissons d'une part de nous focaliser sur ces deux espèces pour cibler nos travaux et les restreindre à une partition des mammifères possédant un système visuel tel que celui décrit en partie 2.1.1. D'autre part, ces deux espèces possèdent des intérêts propres et des complémentarités qui les rendent particulièrement pertinentes pour la problématique de la thèse.

Ainsi, les humains sont une espèce largement étudiée sous des aspects qui éclairent la question de l'attractivité des visages autant en neurosciences, en psychologie qu'en sciences de l'évolution ou en écologie, comme nous l'avons montré dans les parties précédentes. Prélever des données expérimentales avec des humains est plus simple qu'avec d'autres animaux puisqu'ils peuvent verbaliser leurs préférences, comme le montre l'existence de bases de données de stimuli visuels évalués [63, 124] (Cf. Fig. 5.1, (Cf. Fig. 5.2)).

Néanmoins, l'étude des comportements sociaux-sexuels chez les humains peut être limitée par le contexte anthropologique spécifique de cette espèce et en particulier par le chevauchement entre des causes d'origine évolutive et des causes culturelles.

Les mandrills sont certes bien moins étudiés que les humains, mais il y a tout de même des travaux sur leurs comportements sociaux [125] et en particulier sur la perception de leurs visages [126, 127]. En effet, une base de données de plusieurs dizaines de milliers de portraits de mandrills a été récemment mise en place dans le cadre d'un projet de suivi de long terme d'une population naturelle (voire partie 5.2.3). Il reste néanmoins plus difficile et contraignant de réaliser des protocoles expérimentaux avec des primates non humains en général et avec des mandrills en particulier, qu'avec des humains, même si des travaux en ce sens ont déjà été réalisés [116, 118]. Enfin les mâles mandrills présentent des caractéristiques physiques, telles que la coloration du visage et du périnée, qui sont absentes ou moins exagérées chez les femelles [17]. Cette importante différence faciale permet de mettre en évidence des prototypes très caractéristiques des mâles et des femelles mandrills. Cela peut notamment s'expliquer par la compétition induite par le régime d'appariement polygyne des mandrills, comme expliqué en partie 2.3.2. Des conséquences socio-sexuelles de cette différence, spécifiquement au niveau des visages, ont été montrées dans des travaux récents [128].

5.1	Force et complémentarités des modèles humains et mandrills . . . . .	35
5.2	Les mandrills : choix de partenaire, signaux visuels et populations étudiées . . . . .	37
5.2.1	Écologie et organisation sociale des mandrills .	37
5.2.2	Signaux visuels, perception des visages et choix de partenaire chez les mandrills . . . . .	37
5.2.3	Le Projet Mandrillus, la base de données de visages de mandrills et les travaux de vision par ordinateur chez les mandrills . . . . .	37
5.2.4	Des mandrills en semi-captivité dans un centre de recherche au Gabon	40

[63] : LIANG et al. (2018), « SCUT-FBP5500 »

[124] : MA et al. (2021), « Chicago Face Database »

[124] : MA et al. (2021), « Chicago Face Database »

[63] : LIANG et al. (2018), « SCUT-FBP5500 »

[125] : SETCHELL et al. (2005), « Sexual Selection and Reproductive Careers in Mandrills (*Mandrillus sphinx*) »

[126] : CHARPENTIER et al. (2022), « Mandrill mothers associate with infants who look like their own offspring using phenotype matching »

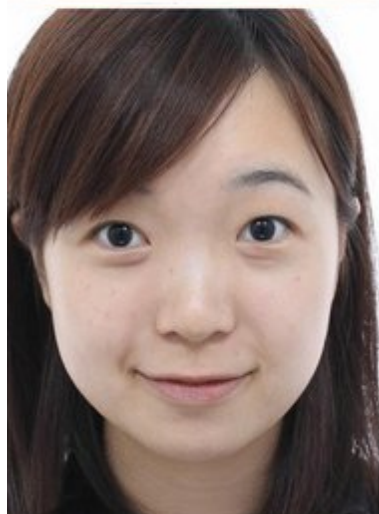
[127] : SETCHELL et al. (2006), « Signal content of red facial coloration in female mandrills (*Mandrillus sphinx*) »

[116] : ROSENFELD et al. (2019), « Experimental evidence that female rhesus macaques (*Macaca mulatta*) perceive variation in male facial masculinity »

[118] : WINTERS et al. (2019), *The structure of species discrimination signals across a primate radiation*



**FIG. 5.1 :** images issues des la base de données de portraits d'humains standardisé CFD (Chicago Face Database) (d'après [124])



**FIG. 5.2 :** images issues des la base de données de portraits d'humains NON standardisé SCUT-FBP5500 (d'après [63])

## 5.2 Les mandrills : choix de partenaire, signaux visuels et populations étudiées

### 5.2.1 Ecologie et organisation sociale des mandrills

Les mandrills (Cf. Fig. 5.3, 5.4, 5.5) sont des primates de la famille des cercopithèques (*Cercopithecidae*), classés comme vulnérables par l'UICN. Ils vivent dans les forêts tropicales d'Afrique centrale, principalement au Gabon, Guinée Equatoriale, dans le sud du Congo et du Cameroun, en groupes allant de plusieurs dizaines à plusieurs centaines d'individus [129], multi mâles et multi femelles. Les femelles restent dans le même groupe de mère en fille (matrilineées) tandis que les mâles migrent dans un autre groupe à l'adolescence, ce qui permet d'éviter les accouplements consanguins [130]. Au sein d'un groupe social, la hiérarchie entre individus est strictement linéaire, et alors que chez les femelles, le rang est principalement transmis de mère en fille, chez les mâles, il peut évoluer tout au long de la vie des individus. La polygynie des mâles implique chez eux une forte compétition intrasexuelle, et cette pression de sélection a pour conséquence un dimorphisme sexuel particulièrement important [125]. En particulier, les mâles sont deux à trois fois plus gros que les femelles (environ 30 kg pour environ 12 kg) et leurs canines mesurent 44mm contre 10mm pour les femelles [131]. Nous allons voir que ce dimorphisme est aussi caractérisé par les traits faciaux des mandrills.

### 5.2.2 Signaux visuels, perception des visages et choix de partenaire chez les mandrills

Darwin affirmait qu' "aucun autre membre de l'entière classe des mammifères n'est coloré, d'une manière aussi extraordinaire que le mandrill mâle" (Cf. Fig. 5.6) [132]. En effet, les mâles adultes présentent des ornements sexuels secondaires extravagants, comme une peau de couleur bleue et rouge sur le visage, la croupe et les organes génitaux, ainsi qu'une barbe jaune, une frange de poils blancs au niveau de l'estomac [133].

La présence de ces signaux sexuels secondaires extravagants chez les mandrills ne serait pas uniquement liée à leur régime d'accouplement polygynie, mais potentiellement aussi aux longs déplacements à travers la forêt et à la grande taille des groupes impliquant que les individus se croisent moins et ont donc besoin de signaux pour se reconnaître [129].

De plus, les mâles choisissent les femelles en fonction de leur statut social, en privilégiant les rangs élevés et préfèrent aussi les femelles qui ont déjà donné naissance [133].

### 5.2.3 Le Projet Mandrillus, la base de données de visages de mandrills et les travaux de vision par ordinateur chez les mandrills

Des mandrills en captivité ont été relâchés en 2002 (36 individus) et en 2006 (29 individus) dans la forêt équatoriale du parc de la Lékédi dans le Haut-Ogooué au sud-est du Gabon [134]. A partir de 2003, des mâles sauvages ont rejoint le groupe pour se reproduire avec les femelles relâchées. Cela

[17] : SETCHELL et al. (2005), « Dominance, Status Signals and Coloration in Male Mandrills (*Mandrillus sphinx*) »

[128] : TIEO et al. (2023), « Social and sexual consequences of facial femininity in a non-human primate »

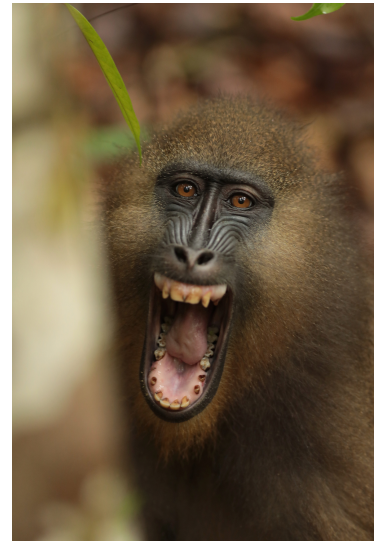


FIG. 5.3 : Image d'une femelle mandrill dans son environnement naturel (Photo prise par Loïc Sauvadet)



FIG. 5.4 : Image d'un mâle mandrill dans son environnement naturel (Photo prise par Loïc Sauvadet)

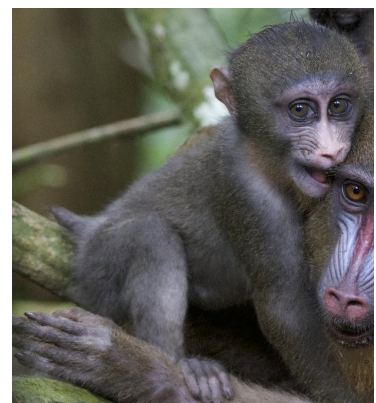


FIG. 5.5 : Image d'un juvénile mandrill dans son environnement naturel (Photo prise par Loïc Sauvadet)

*leucophæus*) the females and young are much paler-coloured, with less green, than the adult males. No other member of the whole class of mammals is coloured in so extraordinary a manner as the adult male mandrill (*Cynocephalus mormon*). The face at this age becomes of a fine blue, with the ridge and tip of the nose of the most brilliant red. According to some authors the face is also marked with whitish stripes, and is shaded in parts



Fig. 67. Head of male Mandrill (from Gervais, 'Hist. Nat. des Mammifères').

with black, but the colours appear to be variable. On the forehead there is a crest of hair, and on the chin a

FIG. 5.6 : Extrait d'une description de Darwin des caractéristiques des mandrills mâles (d'après [132])

[129] : ABERNETHY et al. (2002), « Hordes of mandrills (*Mandrillus sphinx*) »

[130] : SETCHELL (2003), « Behavioural Development in Male Mandrills (*Mandrillus sphinx*) »

[125] : SETCHELL et al. (2005), « Sexual Selection and Reproductive Careers in Mandrills (*Mandrillus sphinx*) »

[131] : SETCHELL et al. (2001), « Changes in the secondary sexual adornments of male mandrills (*Mandrillus sphinx*) are associated with gain and loss of alpha status »

[132] : DARWIN et al. (1981), *The Descent of Man, and Selection in Relation to Sex*

permet un suivi en milieu naturel de ce groupe d'individus, qui est le seul habitué à la présence humaine au monde. Le Projet *Mandrillus* a été mis en place en 2012 pour coordonner ce suivi autour d'une équipe de chercheuses et chercheurs, assistantes et assistants de terrains, étudiantes et étudiants, à la fois internationaux et Gabonais. Cette équipe suit le groupe au quotidien et prélève des données comportementales (focales), biologiques (prises de sang, échantillons fécaux) et des photographies, en particuliers des portraits des individus. A l'heure actuelle, le groupe comporte plus de 350 individus dont la plupart sont nés dans la nature.

Depuis le début du projet, une base de données de ces photographies, la *Mandrillus Face Database (MFD)* [64] a été mise en place. Elle contient<sup>1</sup> 29495 portraits de 397 individus différents (191 femelles, 203 mâles, 3

inconnus). Chaque image est cadrée autour de la tête de l'individu, redressée, et annotée (Cf. Fig. 5.7, 5.8, 5.9). Les variables descriptives des images sont la date de la photo, l'identifiant de l'individu, son sexe, sa date de naissance, ainsi que deux variables décrivant la qualité de l'image, et l'orientation de la tête de l'individu.

Dès 2011, des approches quantitatives basées sur les mesures d'un spectrophotomètre ont été utilisées pour évaluer la relation entre les couleurs du visage des mandrills et le statut social des individus [135]. La création de la MFD a permis de mettre en place des travaux sur les caractéristiques des visages d'une toute autre ampleur pour tester des hypothèses biologiques à propos des mandrills avec des méthodes d'apprentissage profond. Dans le domaine de la sélection de parentèle<sup>2</sup>, à partir de mesures de ressemblance dans un espace latent d'encodage, il a été montré que les apparentés paternels se ressemblent plus que les apparentés maternels [126], et que les mères font en sorte d'orienter les opportunités sociales de leur progéniture vers d'autres enfants qui leurs ressemblent. Dans le domaine de la sélection sexuelle, des travaux basés sur une approche similaire ont mis en évidence des relations entre attractivité et féminité chez les femelles mandrills : les mâles préféreraient des femelles moins féminines [128].

[132] : DARWIN et al. (1981), *The Descent of Man, and Selection in Relation to Sex*

[133] : SETCHELL et al. (2006), « Mate Choice in Male Mandrills (*Mandrillus sphinx*) »

[129] : ABERNETHY et al. (2002), « Hordes of mandrills (*Mandrillus sphinx*) »

[133] : SETCHELL et al. (2006), « Mate Choice in Male Mandrills (*Mandrillus sphinx*) »

[134] : PEIGNOT et al. (2008), « Learning from the first release project of captive-bred mandrills *Mandrillus sphinx* in Gabon »

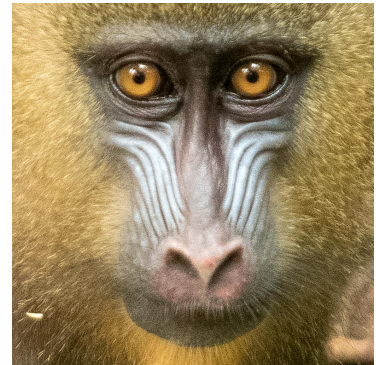
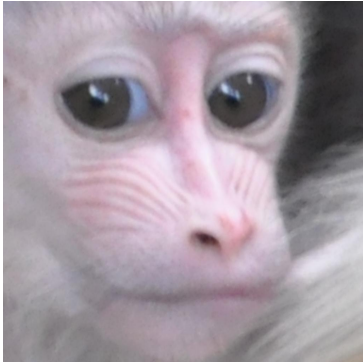


FIG. 5.7 : Portrait d'une femelle mandrill standardisée de la Mandrillus Face Database



FIG. 5.8 : Portrait d'un mâle mandrill standardisé de la Mandrillus Face Database





**FIG. 5.9 : Portrait d'un juvénile mandrill standardisé de la Mandrillus Face Database**

[64] : TIEO et al. (2023), « The Mandrillus Face Database »

1 : pour la version utilisée dans le cadre de cette thèse

[135] : RENOULT et al. (2011), « The Evolution of the Multicoloured Face of Mandrills »

2 : branche de la sélection naturelle expliquant les comportements altruistes envers les individus apparentés

[126] : CHARPENTIER et al. (2022), « Mandrill mothers associate with infants who look like their own offspring using phenotype matching »

[128] : TIEO et al. (2023), « Social and sexual consequences of facial femininity in a non-human primate »



**FIG. 5.10 : Photographie d'une femelle mandrill portant un juvénile, captifs dans leur enclos**

## 5.2.4 Des mandrills en semi-captivité dans un centre de recherche au Gabon

Une autre population de mandrills sur laquelle sont basés de nombreux travaux scientifiques est celle du centre de primatologie (CDP) du Centre Interdisciplinaire de Recherches Médicales de Franceville, aussi situé dans le Haut-Ogooué au Gabon. Il s'agit de mandrills semi-captifs dans des enclos de plusieurs hectares contenant un environnement de type forêt tropicale (Cf. Fig. 5.10).

Cette population a été créée en 1983 avec 15 mandrills des deux sexes dans un premier enclos. Un deuxième enclos a ensuite été créé en 1994 en divisant le groupe du premier enclos qui avait augmenté, uniquement par reproduction. Un troisième enclos a été créé plus tard, permettant l'étude de 3 groupes différents et indépendants. Les individus peuvent se nourrir à partir de ce qu'ils trouvent dans les enclos, mais leurs ressources sont principalement fournies par les équipes d'animaliers du CDP qui organisent un nourrissage quotidien dans des cages en périphérie des enclos, en plus d'un système d'eau illimitée [133]. Ce système de cage permet d'habituer les mandrills à se rendre dans des lieux fermés, offrant la possibilité de réaliser des expérimentations ou pratiquer des soins sur les individus. Plus récemment, les groupes ont augmenté en taille suite à des saisies chez des particuliers qui détenaient des mandrills illégalement (communication personnelle)

Par rapport aux mandrills sauvages, les mandrills mâles du CDP ne peuvent pas migrer entre groupes, ce qui induit une compétition intra-sexuelle plus élevée [125]. Ainsi, on observe que les mâles sont encore plus colorés que ceux en pleine nature, mais de plus petite taille. Concernant la taille des groupes, environ 50 individus par enclos, ils correspondent aux plus petits groupes que l'on peut trouver dans la nature.

Cette thèse a eu pour but de caractériser l'attractivité des visages, avec des approches à la fois de modélisation et expérimentales, mais ayant pour point commun le levier de l'apprentissage profond, génératif et prédictif. Nous avons dans les parties précédentes traité ce sujet sous plusieurs angles différents qui ont été choisis pour permettre de comprendre l'articulation des travaux présentés dans ce manuscrit.

Nous nous sommes intéressés aux mécanismes neuronaux de l'attractivité, en particulier au travers des théories du codage efficient et de la fluence, sans négliger le fait que des caractéristiques visuelles faciales peuvent aussi être des indicateurs de qualité. Pour ce faire, nous avons choisi deux modèles d'études, les mandrills et les humains, nous permettant d'explorer des aspects complémentaires de cette attractivité faciale.

Les 3 chapitres de cette thèse, chacun sous la forme d'un article, proposent des approches différentes pour répondre à la problématique des caractéristiques de l'attractivité des visages, au regard des thèmes qui auront été développés dans l'introduction générale.

Le premier chapitre propose une approche basée sur un développement récent et performant de l'IA générative d'images : StyleGAN3 (Cf. partie 4.2.2), appliqué à la féminité des mandrills. Cette féminité peut être vue comme un cas particulier de prototypicalité (Cf. parties 2.2.2 et 3.2), mais prend aussi sens comme facteur d'attractivité dans le contexte de l'espèce hautement dimorphique que sont les mandrills (Cf. parties 2.3.2, 5.1 et 5.2.1). Ici, au-delà de simplement appliquer des outils existants à de nouvelles images, nous proposons une approche permettant de les synthétiser, de les éditer selon un axe de féminité, mais aussi de contrôler plus finement leur féminité au regard d'une distribution réelle. Ainsi, en nous appuyant sur les données de terrain d'un projet de suivi à long terme, et en particulier sur une base d'images de plusieurs milliers de portraits de mandrills (Cf. partie 5.2.3), nous proposons une méthode innovante qui ouvre de nouvelles perspectives en écologie visuelle expérimentale. L'article qui a émergé de ces travaux a été soumis à *ACM Transactions on Multimedia Computing, Communications, and Applications* et est en attente de réponses des relecteurs.

Le second chapitre s'inscrit à la fois dans la continuité de la méthode développée dans le premier chapitre, et de résultats d'une étude corrélative [64] sur une population de mandrills sauvages (Cf. 5.2.2). Nous proposons de tester expérimentalement ces résultats montrant que les femelles les moins féminines sont préférées par les mâles que chez les mandrills, à l'inverse de ce qui est observé chez les humains. Ainsi, dans une population de mandrills en semi-captivité habituée aux expérimentations en conditions contrôlées (Cf. partie 5.2.4), nous confirmons et renforçons les résultats obtenus précédemment grâce à une expérience basée sur un test de choix et le paradigme du temps d'observation (Cf. partie 4.1.3). L'article produit sur les résultats de ces travaux est en préparation.

Enfin, au-delà de fabriquer des stimuli, l'IA peut aussi servir de modèle de perception (Cf. partie 3.2). En particulier, nous utilisons dans ce troisième

[64] : TIEO et al. (2023), « The Mandrillus Face Database »

chapitre un réseau convolutif (Cf. partie 3.1.2) comme modèle du système perceptif des mammifères (Cf. partie 2.1.1) pour évaluer le lien entre un proxy de la fluence neuronale, la sparsité (Cf. partie 3.2), et la perception de la beauté et de l'attractivité de visages et d'œuvres d'arts par des sujets humains. L'article est publié dans PLOS Computational Biology [136].

[136] : DIBOT et al. (2023), « Sparsity in an artificial neural network predicts beauty »

**CHAPITRE 1 : GENERATION AND  
EDITING OF MANDRILL FACES :  
APPLICATION TO SEX EDITING AND  
ASSESSMENT**



Ce premier chapitre, sous la forme d'un article soumis à *ACM Transactions on Multimedia Computing, Communications, and Applications*, propose un développement méthodologique basé sur l'IA générative permettant de générer, modifier et évaluer quantitativement la féminité faciale de portraits de mandrills.

Cette approche pose les bases techniques pour mettre en place une expérience comportementale d'étude des préférences pour la féminité chez les mandrills, que nous développerons dans le chapitre suivant.

# Generation and Editing of Mandrill Faces: Application to Sex Editing and Assessment

NICOLAS DIBOT, CEFE, Univ. Montpellier, CNRS, EPHE, IRD, France

JULIEN RENOULT, CEFE, Univ. Montpellier, CNRS, EPHE, IRD, France

WILLIAM PUECH, LIRMM, Université de Montpellier, CNRS, France

Generative AI has seen major developments in recent years, enhancing the realism of synthetic images, also known as computer-generated images. In addition, generative AI has also made it possible to modify specific image characteristics through image editing. Previous work has developed methods based on generative adversarial networks (GAN) for generating realistic images, in particular faces, but also to modify specific features. However, this work has never been applied to specific animal species. Moreover, the assessment of the results has been generally done subjectively, rather than quantitatively. In this paper, we propose an approach based on methods for generating images of faces of male or female mandrills, a non-human primate. The proposed method is also capable of editing their sex by identifying a sex axis in the latent space of a specific GAN. Statistical features extracted from real image distributions have been used to develop the assessments. The experimental results we obtained from a specific database are not only realistic, but also accurate, meeting a need for future work in behavioral experiments with wild mandrills.

CCS Concepts: • **Applied computing**; • **Computing methodologies** → **Machine learning algorithms**;

Additional Key Words and Phrases: GenAI, deep learning, image editing, assessment, StyleGAN3, synthetic images, primate mate choice, behavioral experiments, visual ecology

## ACM Reference Format:

Nicolas Dibot, Julien Renoult, and William Puech. 2024. Generation and Editing of Mandrill Faces: Application to Sex Editing and Assessment. *J. ACM* 37, 4, Article 111 (August 2024), 23 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

For more than a decade, AI and more specifically machine learning (ML) has revolutionized a wide range of fields, and image processing in particular. With deep learning, particularly with CNNs, which were revived by LeCun *et al.* in 2015 for image classification [23], one of the most common ML tasks involves training a model to predict labels for input data. However, ML encompasses many other tasks such as; segmentation [26], clustering [6], and, regression and quality assessment [45]. The role of these ML systems is to learn patterns and relationships from training data, using a variety of techniques. They require the creation of mathematical models capable of generalizing from training data such as images to make predictions on new data.

---

Authors' Contact Information: Nicolas Dibot, CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Montpellier, France, nicolas.dibot@cefe.cnrs.fr; Julien Renoult, CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Montpellier, France, julien.renoult@cefe.cnrs.fr; William Puech, LIRMM, Université de Montpellier, CNRS, Montpellier, France, william.puech@lirmm.fr.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

53 More recently, these ML models have been used to generate synthetic images, also known as computer-generated  
54 images, and perform various tasks on them [5]. These models, known as generative AI (GenAI), and, in particular,  
55 generative adversarial networks (GAN), have demonstrated their performance [11]. One of the most popular tasks at  
56 the moment is style transfer, which consists of editing an image to transform it into another image with a different  
57 style [24, 25]. Among the most popular GenAI models, the StyleGAN3 architecture has proved particularly innovative  
58 in combining GAN with the style transfer technique [19]. Innovations have even made it possible to use StyleGAN to  
59 modify real images according to variables of interest.  
60  
61

62 In the fields of ecology and evolution, deep learning has already been used for detection [42, 48] or classification  
63 tasks [40]. Previous GenAI-based approaches have been proposed in ecology [14, 32, 36], but not specifically for animal  
64 behavior, including human behavior, or for the study of visual signals. However, work on mandrill behavior has been  
65 carried out using traditional computer vision metrics [29] or non-generative deep learning [40].  
66

67 In this paper, we propose a method for generating artificial mandrill faces, a non-human primate from central Africa,  
68 and then editing them according to a specific characteristic which is the sex level, for mandrill males (masculinity) and  
69 females (femininity). Furthermore, in our approach, we propose to assess this sex-level variation as a function of a real  
70 distribution. In order to confirm the realism and the visual quality of the images of mandrill faces obtained, we applied  
71 our approach to a specific database containing a large number of mandrill faces from Gabon, both male and female, the  
72 Mandrillus Face Database (MFD) [41].  
73  
74

75 The main contributions we propose in this paper are:

- 76 • An application of GAN-based synthetic image generation to face images of a primate species, the mandrill;
- 77 • Editing the sex level of these mandrill faces;
- 78 • A method for quantitative assessment of the variable used for editing, with a perspective of application to  
79 experimental choice tests for a behavior analysis.  
80  
81

82 The rest of this paper is organized as follows. In Section 2, we first detail current state-of-the-art approaches, related  
83 to GAN and image editing, and then present previous studies on animal behavior and ecology studies using deep  
84 learning and generative AI. Section 3 describes the proposed method for editing mandrill faces and assessing their sex  
85 level. Section 4 presents experimental results obtained with the method we propose, followed by a discussion. Finally,  
86 this paper is concluded in Section 5.  
87  
88

## 90 2 Related work

91 In this section, we present the previous work on which our proposed method is based. In Section 2.1, we first detail  
92 some generative AI concepts, specifically GAN. The application of these concepts to image editing is then discussed in  
93 Section 2.2. Finally, links with the field of ecology, and more specifically with the study of animal behavior are then  
94 detailed in Section 2.3.  
95  
96

### 98 2.1 Generative AI, GAN and encoding

99 Generative adversarial networks (GAN) were first developed 10 years ago [11]. They are based on an adversarial training  
100 mechanism where two neural networks, a generator and a discriminator, are trained together. While the generator  
101 aims to create realistic synthetic images whose feature distribution matches that of a real training image database,  
102  
103



105 the discriminator must learn to distinguish between images from the same real image database and the generator's  
106 synthetic images. Subsequently, major improvements were proposed by the Wasserstein GAN [4] which improves the  
107 convergence and the realism of these models, using an optimization strategy based on optimal transport [43], while  
108 the Progressive Growing GAN generates more realistic images [18]. Afterwards, StyleGAN significantly improves the  
109 realism of the images, particularly portraits [20]. This architecture was enhanced in 2020 with StyleGAN2 [21] and in  
110 2021 with StyleGAN3 [19]. StyleGAN3 is particularly powerful and realistic, enabling the construction of latent spaces  
111 where image features are linear, that is, disentangled [5, 46]. The specificity of the family of StyleGAN models lies in the  
112 integration of a third neural network, called the mapping network, in addition to the generator and discriminator. This  
113 additional network generates a feature vector, which is injected into the generator at various stages, enabling fine-tuned  
114 control of the attributes of the generated image. In comparison, to date and to our knowledge, there are no similar  
115 approaches based on other generative deep learning architectures for editing, encoding and precisely manipulating  
116 images at the same time. Results obtained by variational autoencoders (VAE) [22] are not as realistic, while diffusion  
117 models although realistic, do not allow easy manipulation of their latent encoding spaces [34].

121 With VAE, by construction, it is trivial to find the position of a given image through its latent vector in the latent  
122 space of the trained model, since these models consist of an encoder and a decoder [22]. GAN, on the other hand, only  
123 have a decoder, but no encoder. They are therefore able to generate realistic synthetic images, but are unable to find  
124 the position of a real image in their latent space. For this purpose, encoder architectures additional to a StyleGAN3  
125 generator have been proposed [5]. In particular, the pSp framework enables fairly faithful encoding of real images [30].  
126 To achieve this, it encodes not in the latent space located at the beginning of the generator, but in a larger space  
127 corresponding to all layers of the generator, this enables greater precision. However, the method is still not perfected  
128 but this is an area for future improvements.

## 133 2.2 Handmade versus GAN-based editing

135 Image editing consists of modifying specific characteristics of an image. For example, it is possible to edit the sex  
136 of a portrait, making it more masculine or feminine, or the age, making it younger or older. Morphing techniques  
137 based on Delaunay triangulation [38] or wavelets [27, 37] are not visually efficient. GAN in general and StyleGAN  
138 in particular offer advantages over these previous methods. GAN latent spaces perform well to predict continuous  
139 variables independently of their generative capacity [28]. In particular, StyleGAN performs well not only to construct a  
140 latent space where a specific variable can be identified, but also to generate images modified according to this specific  
141 variable [2]. Several methods have been proposed to identify the axis that describes this specific variable in the GAN  
142 latent space. The supervised linear method uses a Support Vector Machine (SVM) to separate two clusters corresponding  
143 to the two extremities of the variable, and to isolate the axis orthogonal to the SVM support vector [3]. In the absence  
144 of labeled data for the supervised linear method, there are also unsupervised linear approaches without labels [17, 44].

## 149 2.3 Deep learning applied to mandrill behavior and visual communication

151 Mandrills, *Mandrillus sphinx*, are a species of primate in the cercopithecidae family [13]. They are classified as vulnerable  
152 by the International Union for Conservation of Nature (IUCN) [35]. Academic work to better understand their behavior  
153 within an evolutionary biology paradigm has been ongoing for many years [7]. More recently, a database of mandrill  
154 portraits has been created, the Mandrillus Face Database (MFD) [41]. The MFD database detailed in Section 4.1, makes

it possible to use computer vision approaches based on deep learning to study mandrill behavior [8, 9, 40]. The use of deep learning in behavioral ecology has increased in recent years to understand the evolution of visual signals and visual perception in many species, beyond the application to mandrills [10, 12, 16]. Generative models such as GAN, have been rarely used in ecology, and usually for quite different questions, such as modeling the evolution of the prey/predator relationships [36], predicting species coexistence patterns [14] or modeling the trajectory of birds [32].

To our knowledge, there are no GAN-based approaches for generating visual stimuli that give rise to behavioral experiments. With this objective in sight, we have developed the method presented in this article to meet a need for ecologists interested in animal behavior using mandrills as a study model.

### 3 The proposed method for editing images of mandrill faces

The aim of this work is to edit images of mandrill faces in such a way as to modify their femininity or masculinity. In this paper, we define the concept of sex level as the level of belonging to one of the two sexes on a one-dimensional axis. In this section, we develop in detail the method we propose for editing mandrill faces in order to modify their sex level by editing the corresponding generated images of mandrill faces.

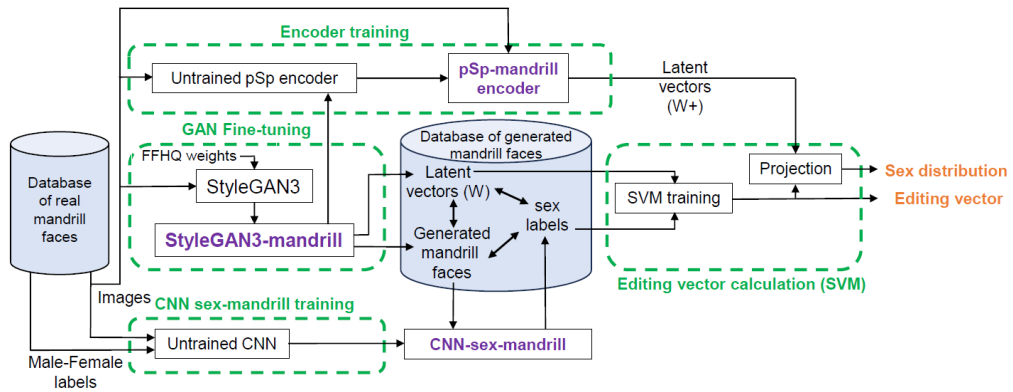


Fig. 1. Overview of the training phase.

The whole process includes a training phase, and an editing phase. The goal of the training phase consists of generating all the necessary processes and data for the editing phase, in particular a trained StyleGAN3 [19] specific for mandrill face images, and a trained pSp encoder [31] for mandrills to generate image latent vector in latent space. The goal of the training phase is also to obtain a sex distribution of mandrill faces and to calculate the editing vector. The objective of the editing phase is to modify the mandrill's masculinity (if it is a male) or femininity (if it is a female) by editing a real image of a mandrill face.

First, in Section 3.1, we give an overview of the training phase, which requires as an input, a database of real images of mandrill faces. The main parts of the training phase are then detailed to explain, in Section 3.2 how to generate a database of mandrill faces from a specific StyleGAN3, in Section 3.3 to train a pSp-mandrill encoder, and in Section 3.4 to determine the editing vector needed for the editing phase. Then, in Section 3.5, we give an overview of the editing

209 phase and we detail in Section 3.6 the computation of the possible edition variation range. Finally, in Section 3.7 we  
210 give details to compute the sex level of an edited generated image of a mandrill face.  
211

### 214 3.1 Overview of the training phase

215  
216 For the editing phase presented in Section 3.5, we first need to train several machine learning algorithms. Indeed, at  
217 the beginning we only have a database of real images of mandrill faces and only untrained or no specific algorithm  
218 architectures. An overview of the training phase is illustrated in Fig. 1.  
219

220 First, in order to generate synthetic images of mandrill faces from a database of real images, we train a StyleGAN3  
221 model [19] on a database of real mandrill faces, the MFD Database. We then obtain a GAN called StyleGAN-mandrill.  
222 During the editing vector computation phase, then we determine in the GAN latent space, noted  $W$ , the vector  $\vec{v}$  that  
223 would feminize or masculinize an image of a mandrill face based on its vector in latent space.  
224

225 This vector is calculated in the space  $W$ . To do this, during the CNN sex-mandrill training phase, we have to also  
226 train a classifier neural network to provide the sex label of the images generated by StyleGAN3-mandrill. Therefore we  
227 have to also train an encoder to find the latent vectors of the real images in the GAN latent space  $W$  in the encoder  
228 training phase. This projection is carried out in a space called  $W+$ . The size of the space  $W+$  is 18 times larger than that  
229 of the space  $W$ . According to the following tasks, we frequently need to switch from the space  $W$  to the space  $W+$ , and  
230 back again. Finally, we project the latent vectors of encoded real images in the space  $W+$  onto the previously identified  
231 vector  $\vec{v}$  to see their distribution for comparison with the femininity or masculinity editing of a single image projected  
232 onto the same axis.  
233  
234  
235  
236  
237

### 238 3.2 Generation of mandrill faces from StyleGAN3-mandrill

239  
240 In order to generate random images, with a size of  $1024 \times 1024$  pixels of mandrill faces, using our approach, we propose  
241 to rely on fine-tuning the StyleGAN3. As illustrated in Fig. 1, to do this, we use a pre-trained version with weights  
242 from the FFHQ (Flickr Faces High Quality) dataset [18], a database of 70,000 human portraits. From these weights, we  
243 fine-tune StyleGAN3 in order to obtain a specific trained StyleGAN3, StyleGAN3-mandrill, by defreezing specific layers  
244 and by using a database of mandrill faces, the Mandrillus Face Database [41], presented in Section 4.1. We evaluate the  
245 training quality using the FID (Frechet Inception Distance) metric [47]. This metric calculates the distance between two  
246 distributions of latent vectors extracted from CNN Inception-V3: one for the real images and the other for generated  
247 images. The obtained results from the FID metric can vary according to the type of data on which the StyleGAN3 is  
248 trained.  
249

250 To generate images with StyleGAN3-mandrill, the model needs to be used as an image generator. Like a decoder, it  
251 transforms small-scale information, *i.e.* a vector of size 512, into large-scale information, *i.e.* a  $1024 \times 1024$  pixel image.  
252 To generate an image with no other purpose than that it be a realistic mandrill face, we simply pass a random latent  
253 vector of a size 512 as input to StyleGAN3-mandrill, which then generates an image of mandrill face. This latent vector,  
254 called  $w$  is injected between each layer of the generator in place of the other latent vector  $w$  generated by the mapping  
255 network when training StyleGAN3.  
256  
257  
258  
259  
260

### 3.3 Training of the pSp-mandrill encoder

Based on the architecture of an untrained encoder pSp [31], and the weights of the trained StyleGAN3-mandrill presented in Section 3.2, the aim of this step is to obtain a pSp encoder trained to specifically encode mandrill faces and called pSp-mandrill encoder as shown in Fig. 1. Editing an image with a GAN to change a feature of interest like sex or age is an algebraic operation. It involves modifying its latent vector  $\mathbf{w}$  along a specific direction. For our approach, this means having its vector  $\mathbf{w}$  in the latent space  $W$  of StyleGAN3-mandrill. When an image is generated by the GAN, it is trivial, because the GAN is a decoder that generates an image from a random latent vector  $\mathbf{w}$  located in the space  $W$  and injected between each layer of the generator. With a real image we can not directly obtain its coordinates. As StyleGAN3 is highly stochastic, it cannot be used “backwards” as an encoder to find the original latent vector precisely. Several previous work propose to overcome this problem [5, 46]. We use the pSp (pixel2style2pixel) encoder [30], which allows us to obtain a latent vector  $\mathbf{w}+$  corresponding to an input image. This latent vector does not correspond solely to the latent vector  $\mathbf{w}$  obtained with StyleGAN3-mandrill (located in the space  $W$ ), but to the concatenation of a set of different latent vectors, which are located in a space called  $W+$ . This set of latent vectors are also injected between each layer of the generator, but whereas in the classical case, the same vector  $\mathbf{w}$  is injected several times, in this case a different vector is injected between each layer. StyleGAN3 containing 18 layers, the dimension of latent vector  $\mathbf{w}+$  is then 18 times larger than the latent latent vector  $\mathbf{w}$ . The proposed approach is the same whether editing in latent space  $W$  or in the latent space  $W+$ .

Since the encoder pSp is also a neural network, it requires training. Its necessary inputs are the images of real mandrill faces and the StyleGAN3-mandrill weights. Starting from scratch, the obtained pSp-mandrill encoder takes a real image of a mandrill face, and provides the latent vector of a real image of mandrill face in the space  $W+$ .

### 3.4 The editing vector calculation

For the editing vector calculation step, from the latent vectors of the generated images, their sex labels and a SVM (Support Vector Machine), we want to calculate the sex vector  $\vec{v}$  giving us the direction along which it is possible to edit femininity or masculinity. In this paper, we postulate that the vector is the same for varying either masculinity and femininity; only the direction of editing changes according to each of the two sexes. The proposed editing vector approach is based on the method proposed by Alaluf *et al.* [33], that we applied to our data.

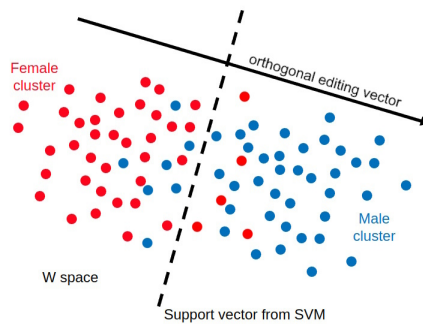


Fig. 2. Calculation of the editing vector from SVM on the clusters of female and male vectors in the space  $W$ .

For this purpose, we first generated 100,000 images of mandrill faces with StyleGAN3-mandrill, to which their latent vectors are associated. Then we determine their sex with a MobileNet architecture CNN classifier [15] trained for this task on sex-labeled images from the database of real mandrill images. Then we train an SVM to classify the latent vectors according to their associated sex. Thus we obtain the vector  $\vec{v}$ , orthogonal to the support vector separating the “male” cluster from the “female” cluster of latent vectors, as illustrated in Fig. 2. Applied to the latent vector of an image as a scalar product, this vector  $\vec{v}$  gives the direction in which to modify a latent vector in order to change the variation of the sex level of the associated image.

The next step is to quantify the variation of the sex level caused by the editing vector  $\vec{v}$  in relation to the distribution of real images, as a ground truth. Then we encode all the real images from the Mandrillus Face Database (MFD) with the pSp-mandrill encoder in space  $W+$ . We project the vectors of these images onto the edit vector  $\vec{v}$ , which we then consider to be a sex axis. This allows us to estimate the distribution of male and female images along the sex axis.

### 3.5 Overview of the editing phase

Now that all the parts are trained, we can integrate them into a pipeline for our mandrill portrait editing application as illustrated in Fig. 3. First we select a real image of a mandrill face. Its sex is determined using the CNN-sex-mandrill classifier presented in Section 3.4. The image is then encoded with the pSp-mandrill encoder in the space  $W+$  and projected onto the sex axis. From its sex level on this axis, we compute a likely editing range corresponding to  $\pm 2$  standard deviations around the mean of the distribution of all real images on this axis. From a given desired sex deviation  $\Delta_d$  of editing, if this is within the range, the latent vector is then edited along the direction given by the sex axis. The new edited latent vector is obtained and decoded with the StyleGAN3-mandrill decoder to obtain a new edited image of a mandrill face.

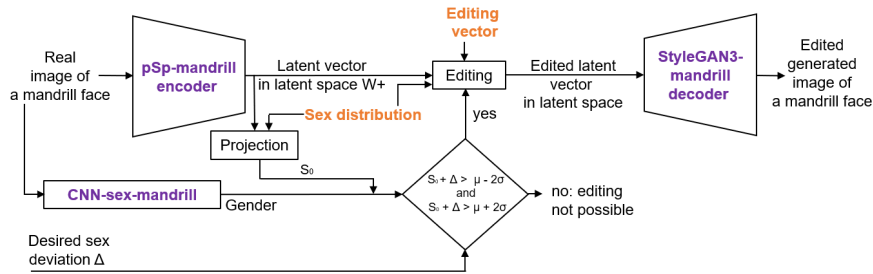
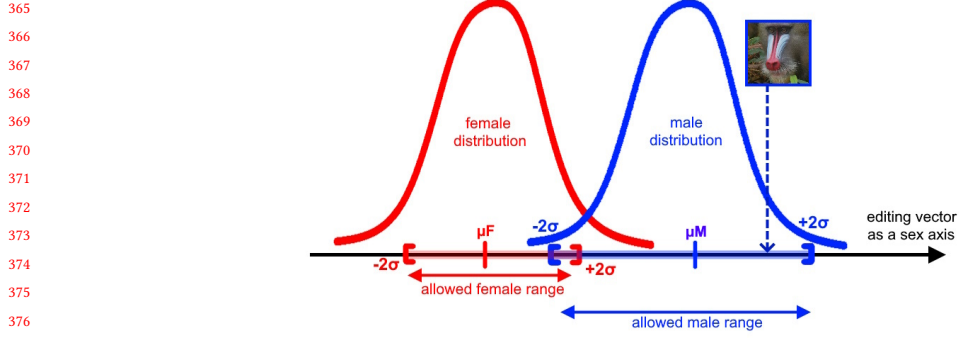


Fig. 3. Overview of the editing phase.

### 3.6 Possibility of the sex level deviation

The editing must modify sex level within a range that exists in reality. In Section 3.4, we explain how to obtain the distribution of real images on the sex axis. We analyze two distributions separately: that of males and that of females. For each, we calculate the mean  $\mu$ , more precisely  $\mu_M$  for males and  $\mu_F$  for females respectively, its standard deviation  $\sigma$ ,  $\sigma_M$  for males and  $\sigma_F$  for females respectively, and determine bounds corresponding to  $\pm 2\sigma$ .



378 Fig. 4. Calculation of the upper and lower edit bounds for each sex according to the distributions of real images projected on the sex axis.  
379  
380

381  
382  
383 Next, we project an image encoded in the space  $W+$  onto the sex axis, and calculate the two distances between its position and the upper and lower bounds of its own sex distribution as shown in Fig. 4. These distances are used to calculate whether or not editing is possible with a desired sex deviation  $\Delta_d$ , as presented in detail in Algorithm 1.  
384  
385  
386  
387  
388

---

389 **Algorithm 1** Calculating whether editing is possible or not.

---

390 1: **Input:** Original image  $I_o$ ; Desired sex deviation  $\Delta_d$   
391 2: **Output:** Boolean *Editing*: Editing possible or not  
392 3:  $Sex \leftarrow \text{CNN}SexMandrill(I_o)$ ; // Sex label classification  
393 4:  $v_o \leftarrow \text{encodeWithSp}(I_o)$ ;  
394 5:  $S_o \leftarrow \text{projection}(v_o)$ ; // Sex level computation  
395 6: **if**  $Sex = \text{Male}$  **then**  
396 7:  $\mu_M \leftarrow \text{averageMale}()$ ;  
397 8:  $\sigma_M \leftarrow \text{standardDeviationMale}()$ ;  
398 9: **if**  $S_o + \Delta_d < \mu_M - 2\sigma_M$  **or**  $S_o + \Delta_d > \mu_M + 2\sigma_M$  **then**  
399 10:  $Editing \leftarrow \text{FALSE}$ ;  
400 11: **else**  
401 12:  $Editing \leftarrow \text{TRUE}$ ;  
402 13: **end if**  
403 14: **else**  
404 15:  $\mu_F \leftarrow \text{averageFemale}()$ ;  
405 16:  $\sigma_F \leftarrow \text{standardDeviationFemale}()$ ;  
406 17: **if**  $S_o + \Delta_d < \mu_F - 2\sigma_F$  **or**  $S_o + \Delta_d > \mu_F + 2\sigma_F$  **then**  
407 18:  $Editing \leftarrow \text{FALSE}$ ;  
408 19: **else**  
409 20:  $Editing \leftarrow \text{TRUE}$ ;  
410 21: **end if**  
411 22: **end if**  
412 23: **return** *Editing*  
413  
414  
415  
416

---

### 3.7 Calculation and assessment of the sex level from the space $W$ to to the space $W+$ of an edited generated image of a mandrill face

In this last step, we explain how to edit and assess the encoded image according to a desired deviation on the sex axis as illustrated in Fig. 3. From the original image  $I_o$ , the desired deviation  $\Delta_d$ , two optimization parameters, the step  $s$  corresponding to the modification intensity of the editing deviation to be optimized, the threshold  $T$  corresponding to the editing precision tolerance, and the standard deviation  $\sigma$ , we want to obtain the edited image  $I_e$  with the optimized editing deviation  $\Delta_e$  that enabled this editing.

In order to edit an image, we cannot directly modify its coordinates according to the latent vector  $v$  multiplied by the desired deviation. In fact, in the space  $W+$ , the magnitude of the editing intensity is not the same depending on where it is measured. It is not the same for an encoded image directly after its edition in the space  $W+$  or for this same image decoded then re-encoded with pSp-mandrill after its edition. As the editing vector is calculated in the space  $W$ , it does not necessarily correspond to the optimal direction of sex level in the space  $W+$ .

To overcome this problem, we have developed an algorithm, presented in Algorithm 2, which optimizes the editing intensity to be applied to the latent vector of an encoded image by decoding and re-encoding the edited image, then comparing the sex level of the re-encoded image with the desired sex level. The principle of Algorithm 2 is to edit an image with a certain deviation  $\Delta$ , see if its sex level corresponds to the desired sex level, then according to this, modify the value of the deviation  $\Delta$  by increasing it, decreasing it, or changing its sign and repeating the same steps until the desired deviation  $\Delta_d$  is reached.

Fig 5 illustrates some of the conditions specified in Algorithm 2. This example, which searches for the optimal editing step  $\Delta_e$  for editing the sex level of a mandrill face, starts from the sex level  $S_0$  of the original generated image, to reach the sex level  $S_R$ . In the first step,  $\Delta_1$  corresponds to condition C of the Alg. 2, for the second step  $\Delta_2$  to condition B, while for the third step  $\Delta_3$  to condition A and finally  $\Delta_4$  to the optimal  $\Delta_e$ . Note that for each step, the algorithm restarts from  $S_0$ .

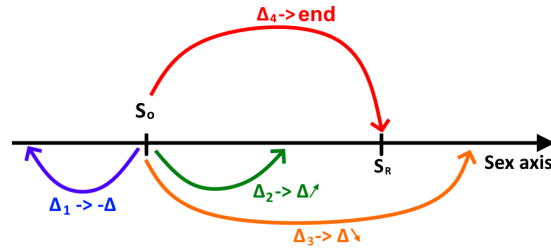


Fig. 5. Example of the research of the optimal editing step  $\Delta_e$  for editing the sex level of a mandrill face. Starting from the sex level  $S_0$  of the original generated image, to reach the sex level  $S_R$ , in the first step,  $\Delta_1$  corresponds to condition C of the Alg. 2, for the second step  $\Delta_2$  to condition B, while for the third step  $\Delta_3$  to condition A and finally  $\Delta_4$  to the optimal  $\Delta_e$ .

## 4 Experimental results

In this section, we present the results obtained with our proposed approach applied to real images of mandrill faces. After presenting the database of mandrill images used in Section 4.1, we apply in Section 4.2 the developed method

**Algorithm 2** Optimal research of the editing step.

---

```

469 Algorithm 2 Optimal research of the editing step.
470
471 1: Input: Image  $I_o$ ; Desired deviation  $\Delta_d$ ;
472 2: Step  $s$ ; Threshold  $T$ ; standard deviation  $\sigma$ ;
473 3: Output: Edited image  $I_e$ ; Optimized deviation  $\Delta_e$ ;
474 4: if editingPossible( $I_o, \Delta_d$ ) then
475 5:   // Original image sex level computation
476 6:    $v_o \leftarrow \text{encodeWithSp}(I_o)$ ;  $S_o \leftarrow \text{projection}(v_o)$ ;
477 7:   // Initialization
478 8:    $S_r \leftarrow S_o + \Delta_d$ ;  $\Delta_e \leftarrow \Delta_d$ ;
479 9:   // Generation of the initial edited image
480 10:   $v_e \leftarrow v_o + \Delta_e \vec{d}$ ;
481 11:   $I_e \leftarrow \text{decodeWithStyleGAN3-mandrill}(v_e)$ ;
482 12:  // Initial edited image sex level computation
483 13:   $v_e \leftarrow \text{encodeWithSp}(I_e)$ ;  $S_e \leftarrow \text{projection}(v_e)$ ;
484 14:  while  $S_e \neq S_r \pm T \times \sigma$  do
485 15:    if  $S_r > S_o$  then
486 16:      if  $S_e > S_o$  then
487 17:        if  $S_e > S_r$  then
488 18:           $\Delta_e \leftarrow \Delta_e - \Delta_e/s$  // Condition A
489 19:        else
490 20:           $\Delta_e \leftarrow \Delta_e + \Delta_e/s$  // Condition B
491 21:        end if
492 22:      else
493 23:         $\Delta_e \leftarrow |\Delta_e| + |\Delta_e|/s$  // Condition C
494 24:      end if
495 25:    else
496 26:      if  $S_e < S_o$  then
497 27:        if  $S_e > S_r$  then
498 28:           $\Delta_e \leftarrow \Delta_e + \Delta_e/s$  // Condition D
499 29:        else
500 30:           $\Delta_e \leftarrow \Delta_e - \Delta_e/s$  // Condition E
501 31:        end if
502 32:      else
503 33:         $\Delta_e \leftarrow -|\Delta_e| + \Delta_e/s$  // Condition F
504 34:      end if
505 35:    end if
506 36:    // Generation of the current edited image
507 37:     $v_e \leftarrow v_o + \Delta_e \vec{d}$ ;
508 38:     $I_e \leftarrow \text{decodeWithStyleGAN3-mandrill}(v_e)$ ;
509 39:    // Current edited image sex level computation
510 40:     $v_e \leftarrow \text{encodeWithSp}(I_e)$ ;  $S_e \leftarrow \text{projection}(v_e)$ ;
511 41:  end while
512 42:  return  $I_e$  and  $\Delta_e$ 
513 43: else
514 44:   return Edition not possible
515 45: end if

```

---

516  
517  
518  
519  
520



to two examples, an image of a male mandrill and an image of a female mandrill, carefully detailing all the steps. We then show the results of the method on various examples in Section 4.3, before discussing some unexpected results and taking a step back from our work in Section 4.4.



Fig. 6. Mandrill #20230104\_id203(2): a) In its environment, b) Cropped and straightened image of its face

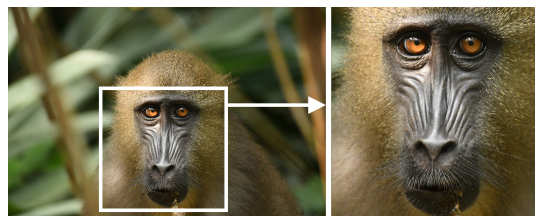


Fig. 7. Mandrill #20230104\_id205(5): a) In its environment, b) Cropped and straightened image of its face.

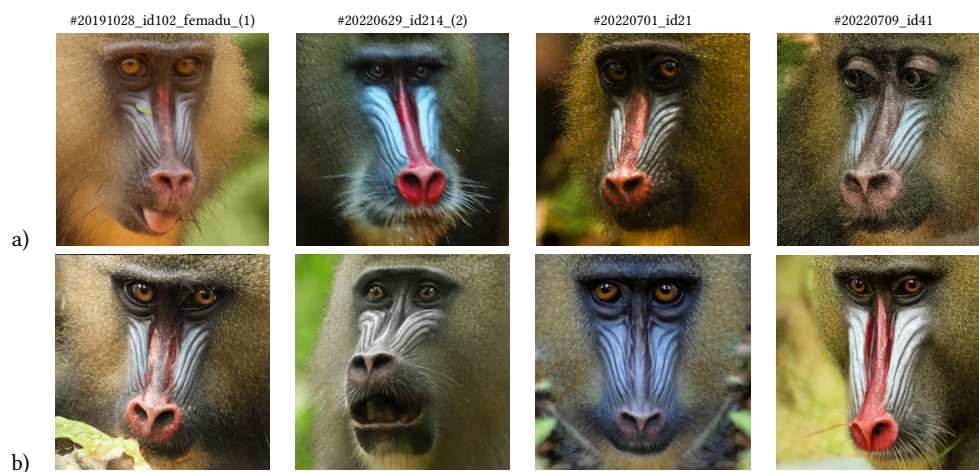


Fig. 8. Comparison between: a) Real mandrill faces from the MFD database [39], and b) Synthetic non-existing mandrill faces generated from StyleGAN3-mandrill

#### 4.1 The Mandrillus Face Database [41]

The Mandrillus Face Database (MFD) [41] is one of the largest non-human animal face databases, as it contains 29495 images, representing 397 various mandrills, from pictures taken over a period of 10 years. The images were taken by volunteers following a group of mandrills through a rainforest in the Lékédi Park and its surroundings, in southern Gabon (near to the village of Bakoumba).

As shown in Fig. 6 and Fig. 7, from a full picture of a mandrill in its natural rainforest environment, the mandrill faces are manually aligned and cropped. Images are first manually oriented to align the pupils of the eyes horizontally, and then centered and cropped to keep only the face, the neck and ears are then removed.

In our work, to train our machine learning algorithms, we use a database subset corresponding to good or high quality images of 4820 adult male and 13538 female mandrill faces, making a total of 18358 images. Good quality and high quality images correspond to “Quality 2” and “Quality 3” images respectively in the MFD database [41].

#### 4.2 Detailed examples, from real mandrill faces to edited generated mandrill faces

In this section, we detail the experimental results we obtained by our proposed method. First, the training step of the developed StyleGAN3-mandrill takes 110 epochs, freezing 11 layers, *i.e.* keeping the pre-trained StyleGAN3 weights on FFHQ (human faces) for the first 11 layers of the GAN generator. Before this training, we have a  $FID = 266.26$ , calculated between synthetic images generated by StyleGAN3 and real images of the MFD database subset. We stop our training when the images generated by the StyleGAN3-mandrill are visually plausible, as shown in Fig. 8, and the FID no longer appears to be decreasing. At the end of training, the FID is 3.65. In Fig. 8.a we show four real mandrill faces from the MFD database [39] while in Fig. 8.b we illustrate four synthetic mandrill faces, that is faces of mandrills that do not exist, generated from StyleGAN3-mandrill. These synthetic images are generated from a random latent vector of a size 512, the number of dimensions of the space  $W$ . These synthetic mandrill faces, Fig. 8.b, have been shown to experts and they were not able to distinguish the synthetic from the real.

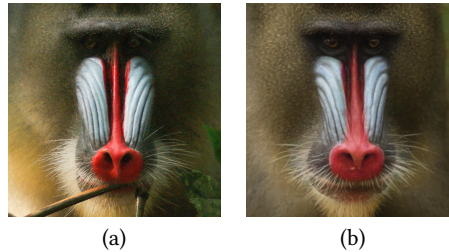


Fig. 9. Male mandrill face #ID-57\_20150517(13) [41]: a) Real image, b) Synthetic image generated from (a) encoded with pSp-mandrill then decoded with StyleGAN3-mandrill, with a sex level  $S_o = 0.59$ .

From a real image of a mandrill face, we can then generate its corresponding synthetic image with an encoding based on pSp-mandrill followed by a decoding with StyleGAN3-mandrill as presented in Section 3.2. Remember, face editing cannot be applied to a real image, but only to a generated one, in order to edit its latent vector. Fig. 9 illustrates the generation of a synthetic mandrill face from an image of a real mandrill face. From the real image of the male mandrill face #ID-57\_20150517(13) [41] shown in Fig. 9.a, we can generate the corresponding synthetic image, shown in Fig. 9.b,

625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676

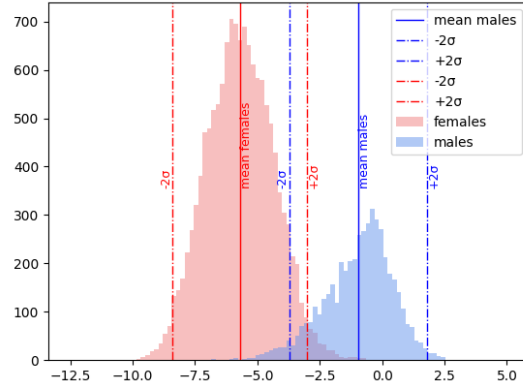


Fig. 10. Distribution of the real mandrill faces from the MFD database subset [39] projected on the sex axis, with the positions of the sex level means ( $\mu_F = -5.70$  and  $\mu_M = -0.96$ ) and the upper and lower bounds (between  $[-8.40, -2.99]$  for females and between  $[-3.72, 1.81]$  for males)

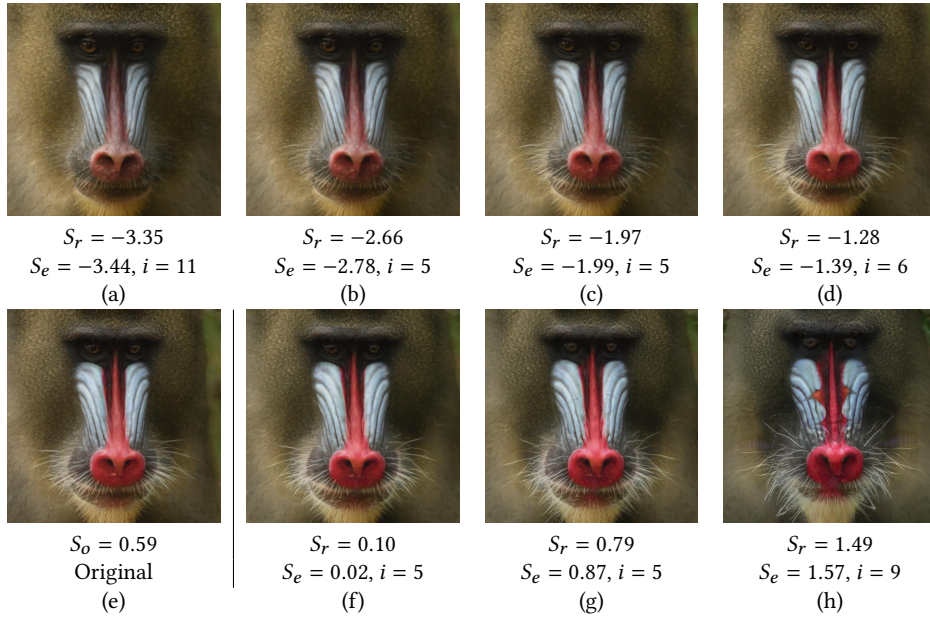
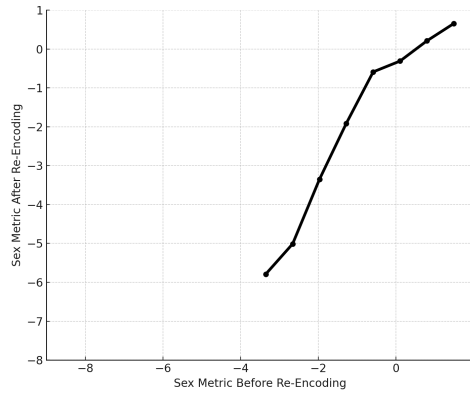


Fig. 11. Editing of the male mandrill face #-57\_20150517(13) [41] from the generated image illustrated Fig. 9.b : a) Edited generated image with a sex level  $S_e = S_o - 2\sigma_M = -3.44$ , b) Edited generated image with a sex level  $S_e = S_o - 1.5\sigma_M = -2.78$ , c) Edited generated image with a sex level  $S_e = S_o - \sigma_M = -1.99$ , d) Edited generated image with a sex level  $S_e = S_o - 0.5\sigma_M = -1.28$ , e) Original encoded image (Fig. 9.b) with a sex level  $S_o = 0.59$ , f) Edited generated image with a sex level  $S_e = S_o + 0.5\sigma_M = 0.02$ , g) Edited generated image with a sex level  $S_e = S_o + \sigma_M = 0.81$ , h) Edited generated image with a sex level  $S_e = S_o + 1.5\sigma_M = 1.57$ . For each image,  $i$  is the number of iterations in Algorithm 2 to obtain the desired deviation  $\Delta_d$ .

677 which has been encoded with pSp-mandrill to obtain its latent vector in the space  $W+$ , with a size 18 times larger than  
 678 that of the space  $W$  (size of 9216), and decoded with StyleGAN3-mandrill to visualize it. In the generated image Fig. 9.b,  
 679 we can see that the branch under the nose of the mandrill disappears during this process, and that colors and shadows  
 680 are slightly modified. This is because the encoder positions the image in the space  $W+$ , according to the distribution of  
 681 features found in the MFD database subset and learned by StyleGAN3-mandrill. As the features of an atypical object  
 682 like a branch are poorly represented in this distribution, such features disappear during the encoding step. Note that  
 683 the decoded image has not been modified in any way. To improve the preservation of shadows and colors, a histogram  
 684 specification step could be implemented. After the projection on the sex axis, the sex level of this generated mandrill  
 685 face image, is determined as sex level  $S_o = 0.59$ .  
 687

688 As shown in Fig. 10, all the images used in the MFD database subset are encoded with pSp-mandrill and projected  
 690 onto the sex axis. This allows us to obtain the distribution of their sex level on this axis. For the projection, as assumed  
 691 in Section 3.4, the sex axis is the same for both males and females. Even with this assumption, we can see on Fig. 10  
 692 that the two distributions are relatively easy to separate, and that on this axis, the sex levels for females are generally  
 693 lower than those for males. Indeed, for the female distribution, from 13538 female mandrill faces, the mean sex level is  
 694  $\mu_F = -5.70$ , with a standard deviation of  $\sigma_F = 1.35$ , while for the male distribution, from 4820 images, the mean sex  
 695 level is  $\mu_M = -0.96$ , with a standard deviation is  $\sigma_M = 1.38$ .  
 697



712  
713 Fig. 12. Sex level editing as a function of the obtained sex levels after decoding and re-encoding for the image #-57\_20150517(13)  
 714 shown Fig. 11.  
 715

716 In the rest of this section we present the results obtained at the different steps of editing real images of mandrill faces,  
 717 with two examples, one of a male and one of a female. Images are then edited at the desired deviation  $\Delta_d$ . In Fig. 11  
 718 we illustrate an example with the male mandrill face generated and illustrated in Fig. 9.b. From this generated male  
 719 mandrill face, illustrated in Fig. 11.e. we choose deviations  $\Delta_e = -2\sigma$ , in Fig. 11.a,  $\Delta_e = -1.5\sigma$ , in Fig. 11.b,  $\Delta_e = -\sigma$ ,  
 720 in Fig. 11.c,  $\Delta_e = -0.5\sigma$ , in Fig. 11.d,  $\Delta_e = 0.5\sigma$ , in Fig. 11.f,  $\Delta_e = \sigma$ , in Fig. 11.g, and  $\Delta_e = 1.5\sigma$ , in Fig. 11.h. Note, that  
 721 based on the conditions presented in Algorithm 1, it is not possible to go as far as  $\Delta_e = -2.5\sigma$  or  $\Delta_e = 2\sigma$  in order not  
 722 to exceed the bounds of  $2\sigma$  around the mean sex level  $\mu_M$ .  
 724

725 To obtain the edited images shown in Fig. 11 with the optimized deviation  $\Delta_e$ , we used Algorithm 2. Experimentally,  
 726 we set the value of the threshold  $T = 0.1$ . This threshold  $T$  is multiplied by the standard deviation in the algorithm,  
 727  
 728

giving a value of 0.138. This value guarantees visual accuracy, knowing that a greater precision would be useless as it would be imperceptible to the naked eye, even for experts. What's more, the relationship between the sex levels of the edited images and the sex levels of the same images decoded and then re-encoded is not monotonic on a scale smaller than  $T = 0.1$ . Consequently, decreasing the value of threshold  $T$  would have no impact on accuracy. We set the value of the step size  $S = 7$ , which means that fewer iterations are needed than with a higher value of  $S$ , while avoiding diverging out of the solution space, which would have been the case with a lower value of  $S$ . For the editing, shown in Fig. 11, the number of iterations of Algorithm 2 varies from 5 to 11 with these parameter values. By using Algorithm 2, the first value of the deviation  $\Delta_e$  before optimization corresponds to the desired editing. There seems to be a monotonic correlation between this edition and the value corresponding to what we obtain after optimization, as can be seen in Fig. 12 with several editing values. In Fig. 12, for the image #-57\_20150517(13) shown in Fig. 11, we can see that for a range between  $-2\sigma$  and  $1.5\sigma$  for the sex level editing, we obtain values for the sex level editing after decoding then re-encoding in a range between  $-6$  and  $1$ . Note that visually, the increased masculinity of an image corresponds to more pronounced features, more contrasting colors, a more elongated face and a redder nose, as shown in Fig. 11.

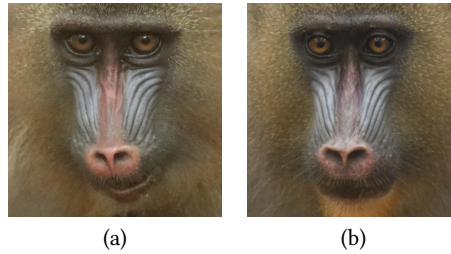


Fig. 13. Female mandrill face #20210325\_id221\_femadu [41]: a) Real image, b) Synthetic image generated from (a) encoded with pSp-mandrill then decoded with StyleGAN3-mandrill, with a sex level  $S_o = -5.70$ .

Similarly, as illustrated in Fig. 13, we have encoded an image of the female mandrill face #20210325\_id221\_femadu, Fig. 13.a, and then encoded with pSp-mandrill then decoded it, Fig. 13.b, in the same way as the male mandrill example. From the real image of the female mandrill face Fig. 13.a, note that the encoding, illustrated Fig. 13.b modifies the inclination of the head and makes the eyes rounder. this is again a result of feature normalization as observed with the example illustrated Fig. 9. 9.

As shown in Fig. 14, the generated female mandrill face Fig. 13.b is also edited with Algorithm 2 according to the corresponding step range for a female. In order not to exceed  $2\sigma$  around the mean sex level  $\mu_F$ , the editing is within a range of  $-2$  to  $+1.5\sigma$ . As with the male mandrill faces, overall, there seems to be a monotonic correlation between this editing and the value corresponding to what we obtain after optimization, as can be seen in Fig. 15 with several editing values. (with a small exception between 0 and  $0.5\sigma$ ). Visually, a more feminine face is rounder in shape, with lighter colors and proportionally larger eyes.

#### 4.3 Additional examples from the MFD database subset

The results presented in Section 4.2 can be applied to any mandrill face image. In this section, we present the pSp-mandrill encoding and editing obtained for 4 images of male mandrill faces, as shown in Fig. 16, and 4 images of female mandrill faces, as shown in Fig. 17, illustrating several examples of positive and negative deviation step size,

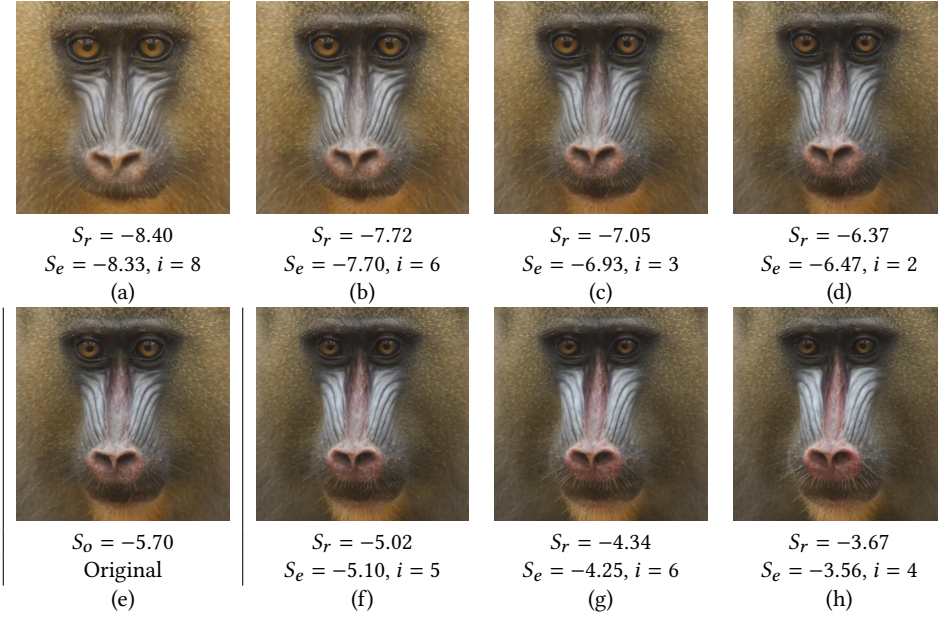


Fig. 14. Editing of the female mandrill face #20210325\_id221\_femadu [41] from the generated image illustrated Fig. 9.b : a) Edited generated image with a sex level  $S_e = S_o - 2\sigma_M = -8.33$ , b) Edited generated image with a sex level  $S_e = S_o - 1.5\sigma_M = -7.72$ , c) Edited generated image with a sex level  $S_e = S_o - \sigma_M = -6.93$ , d) Edited generated image with a sex level  $S_e = S_o - 0.5\sigma_M = -6.47$ , e) Original encoded image (Fig. 13.b) with a sex level  $S_o = -5.70$ , f) Edited generated image with a sex level  $S_e = S_o + 0.5\sigma_M = -5.10$ , g) Edited generated image with a sex level  $S_e = S_o + \sigma_M = -4.25$ , h) Edited generated image with a sex level  $S_e = S_o + 1.5\sigma_M = -3.56$ .

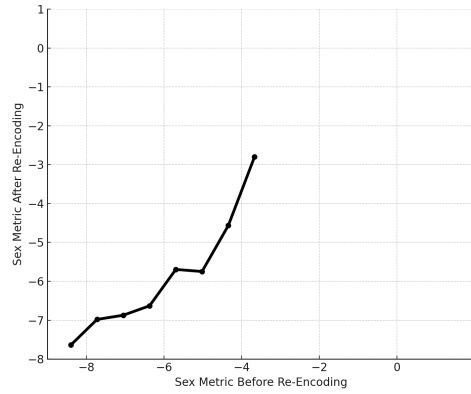
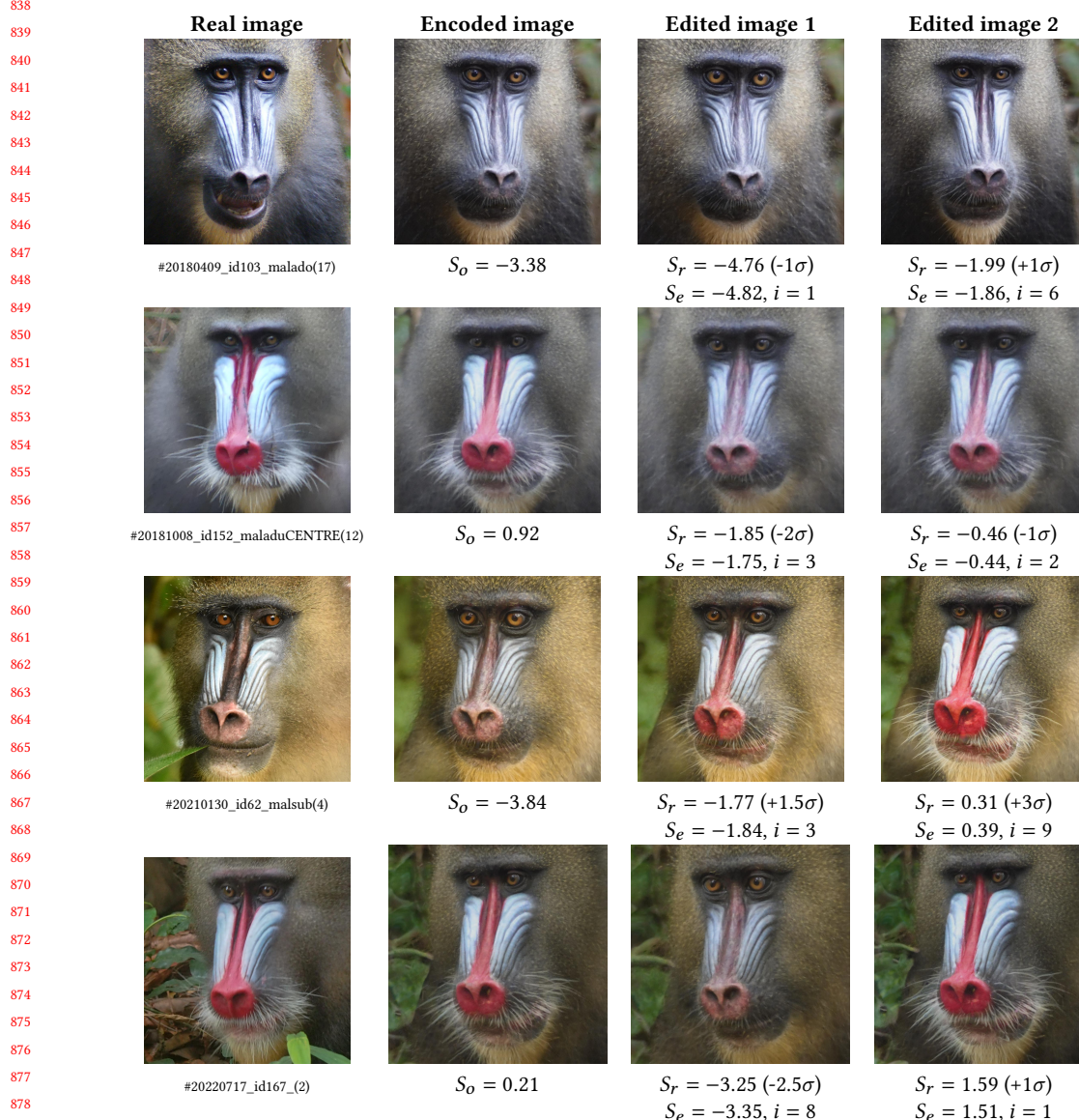


Fig. 15. Sex level editing as a function of the obtained sex levels after decoding and re-encoding for the image #20210325\_id221\_femadu shown Fig. 14.

within the bounds  $\mu \pm 2\sigma$ . Note that an original image can still be edited if its sex level is outside the bounds, as in the case of the image #20210130\_id62\_malsub(4), Fig. 16, whose original sex level of  $-3.84$  is below the minimum male sex level limit of  $\mu - 2\sigma = -3.72$ . In this case, editing can only be done by increasing the sex level. For the image

Manuscript submitted to ACM

833 #20210130\_id62\_malsub(4), Fig. 16, the edited image 1 is obtained by increasing the sex level of  $1.5\sigma$  and the edited  
 834 image 2 of  $3\sigma$ . It can be seen that for editions close to the upper limit of  $+2\sigma$  for females and lower limit of  $-2\sigma$  for males,  
 835 the sex of the mandrills is ambiguous, which is expected as this corresponds to the area of overlap between the two  
 836 distributions on the sex axis.  
 837



880 Fig. 16. Examples of encoded and edited male mandrill face images: The first column is the original image from the MFD database;  
 881 The second column is the image encoded with pSp-mandrill; The third and fourth columns are edited images with different editing  
 882 values, the first being the least masculine and the second the most masculine.  
 883  
 884

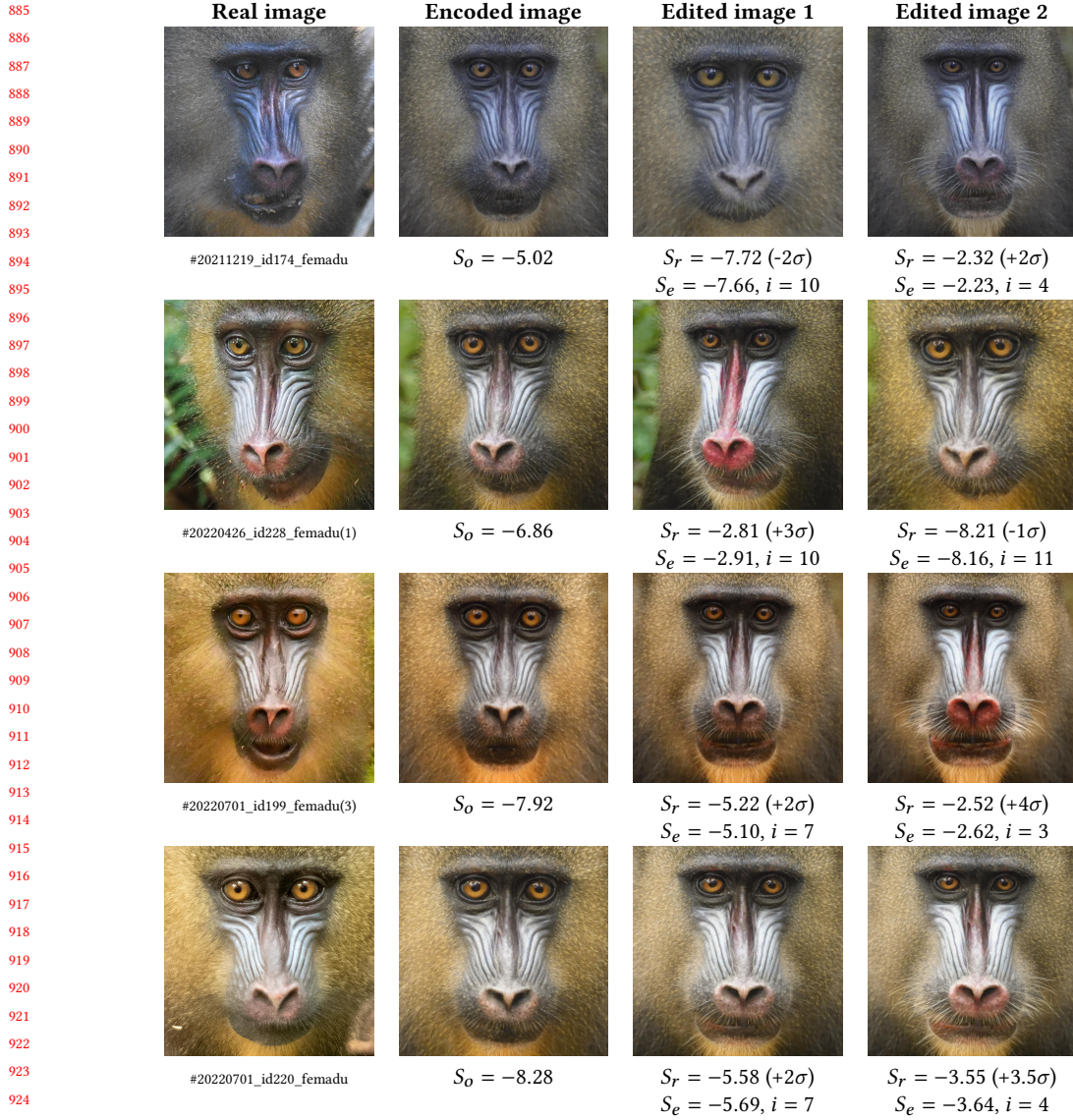


Fig. 17. Examples of encoded and edited female mandrill face images: The first column is the original image from the MFD database; The second column is the image encoded with pSp-mandrill; The third and fourth columns are edited images with different editing values, the first being the most feminine and the second the least feminine.

Fig. 18 shows the relationships of the sex levels before the decoding-re-encoding step as a function of sex levels obtained after the decoding-re-encoding step, for all the editing performed and illustrated in Fig. 11, Fig. 14, Fig. 16 and Fig. 17. We observe an almost monotonic increasing relationship between the two axes, although this is not strictly the case everywhere, which justifies the conditions  $C$  and  $F$  of Algorithm 2, where the  $\Delta_e$  optimization can be decreasing.

Manuscript submitted to ACM



937 However, there is a visual tendency towards a linear or logit relationship between the two axes. Thus, estimating the  
 938 first value of  $\Delta_e$  with a predictive model adapted to these points is an area for improvement to reduce the number  
 939 of iterations of Algorithm 2. However, note that because convergence was reached very quickly (in a few iterations)  
 940 with our approach, it is not sure that such an analytic approach is worthwhile. It can also be seen that females have  
 941 predominantly smaller sex level values than males, which corresponds well to the distribution of real images projected  
 942 on the axis.  
 943  
 944

945  
 946  
 947  
 948  
 949  
 950  
 951  
 952  
 953  
 954  
 955  
 956  
 957  
 958  
 959  
 960  
 961  
 962  
 963  
 964  
 965  
 966  
 967  
 968  
 969  
 970  
 971  
 972  
 973  
 974  
 975  
 976  
 977  
 978  
 979  
 980  
 981  
 982  
 983  
 984  
 985  
 986  
 987  
 988

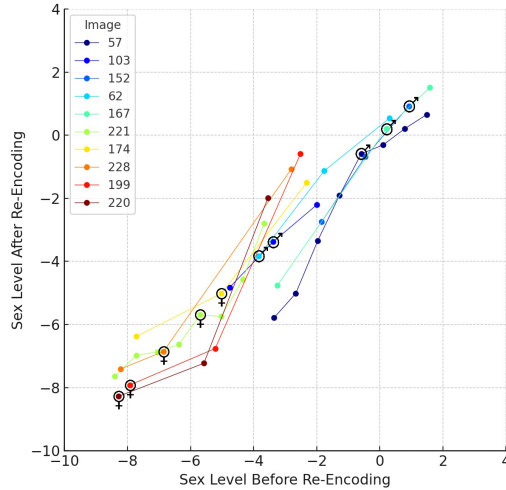


Fig. 18. Sex level editing as a function of the obtained sex levels after decoding and re-encoding for all the editing performed and illustrated Fig. 11, Fig. 14, Fig. 16 and Fig. 17.

#### 4.4 Discussion

In this section, we propose to discuss some of the results obtained, which present some particular aspects. First of all, we have chosen to limit ourselves to bounds of  $+2\sigma$  and  $-2\sigma$  around the sex level averages of each sex when editing the images. Indeed, generative AI models are generally good at interpolating between images in the training database, but poor at extrapolating beyond this data. For example as shown in Fig. 19, with an editing reaching  $+4\sigma$  from the male mean, the image of the edited mandrill face shows artifacts that make it unrealistic. However, experts accustomed to observing mandrills have confirmed the plausibility of these types of editing, when they remain between bounds.

In addition, behavioral and non-visual features seem to emerge from the editing. In some cases, the more masculine an image is edited, the more the mandrill’s mouth is open, as shown in the last column of Fig. 16, in particular for the second and the fourth examples. This may be explained by the fact that male mandrills in the MFD database are more aggressive and open their mouths to scream. In other cases, the more feminine an image is edited, the more the mandrill turns its head away. Indeed, female mandrills are more fearful and turn their heads when a photo is taken to run away.

It is also noticeable that the background of the images behind the mandrills is slightly modified by the editing process. This shows that the StyleGAN3-mandrill’s space  $W+$  does not allow perfect disentanglement of the gender variable.

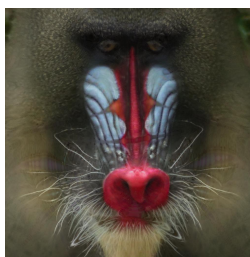


Fig. 19. Editing out of bounds of the face of the mandrill #20150517\_id57\_maladu\_(13), showing some artefacts.

Some work [3] suggests that editing in a different GAN space, called S-space, could improve this, which is a direction for future improvement.

Beyond results, the method itself presents some points to discuss. While the sex axis is calculated in the space  $W$ , editing is performed in the space  $W+$ , duplicating the axis 18 times. This can lead to some inaccuracy, with editing in certain dimensions that do not correspond precisely to the sex, but more to the age, for example. This probably explains the need to optimize the editing level with Algorithm 2. One way of improving this would be to train the SVM directly in the space  $W+$ , by re-encoding in the space  $W+$  the images generated from  $W$  and use them for training.

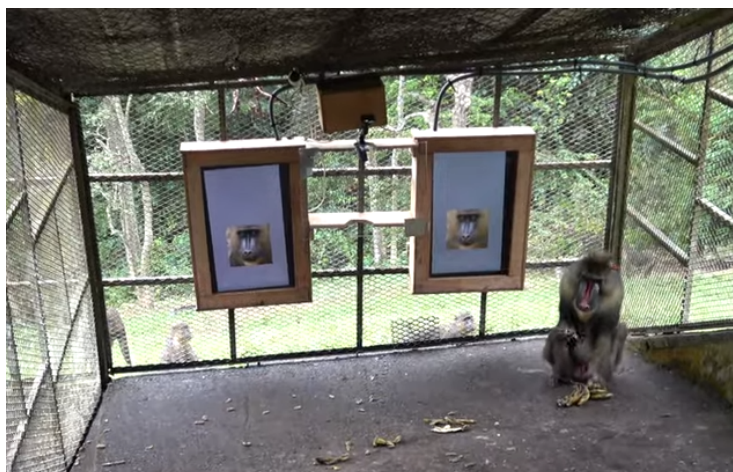


Fig. 20. Illustration showing the use of the proposed methods to conduct behavioral choice experiments with semi-captive mandrills in Gabon. The aim is to determine whether sex level has an influence on mate choice.

The methods proposed in this paper aim to conduct behavioral choice experiments with semi-captive mandrills in Gabon. The aim is to determine whether sex level has an influence on mate choice. To improve the attractiveness of synthetic mandrill images for real mandrills, we animated the images by editing them according to head orientation, thus creating a video of a mandrill moving its head from left to right. As shown in Fig. 20, two images of mandrill faces of the same individual are presented to a mandrill of the opposite sex, on two juxtaposed screens. Using various sensors, including cameras, we then measure which screen attracts the observed mandrill the most. These approaches could be

1041 applied in many different ways. We could go even further by exploiting the potential of GAN to improve the quality of  
1042 the images generated [1].  
1043

## 1044 5 Conclusion

1046 In this paper, we presented a GenAI-based framework for generating, editing and assessing images of mandrill faces  
1047 based on their sex. More precisely, we used a StyleGAN3 trained on a database of thousands of real mandrill images,  
1048 StyleGAN3-mandrill, to generate synthetic images of these primates, coupled with an encoder trained on this same  
1049 database, pSp-mandrill, to encode the real images in a latent space close to that of the trained GAN. We succeeded  
1050 in finding a sex axis on these two spaces ( $W$  and  $W+$ ) to edit the images in order to change their sex level and thus,  
1051 make them more or less feminine or masculine. In addition, we proposed a method for assessing this editing through  
1052 statistics of the distribution of real images in the database on the sex axis. Our approach not only shows the potential of  
1053 the StyleGAN3-based framework in animal image manipulation, but also helps scientists working with visual stimuli to  
1054 design stimuli with controlled variation for psychological and behavioral experiments with both humans and animals.  
1055

1058 This work has also revealed several limitations of the GenAI-based approach, such as an imperfect disentanglement  
1059 of the features of interest, which will be the focus of features research. This work would also be enhanced if applied  
1060 to other animal species, included humans, which would nevertheless require a database on the same scale as that for  
1061 mandrills.  
1062

## 1064 Acknowledgments

1066 This work was supported in part by the Agence Nationale de la Recherche (ANR-20-CE02-0005-01), in part by the CNRS  
1067 through the MITI interdisciplinary programs (Programme Interne Blanc MITI 2023.1 - Projet: DEEPCOM- L'intelligence  
1068 artificielle pour étudier la communication).  
1069

## 1071 References

- 1072 [1] Lorenzo Agnolucci, Leonardo Galteri, Marco Bertini, and Alberto Del Bimbo. 2024. Perceptual Quality Improvement in Videoconferencing Using  
1073 Keyframes-Based GAN. *IEEE Transactions on Multimedia* 26 (2024), 339–352. <https://doi.org/10.1109/TMM.2023.3264882>
- 1074 [2] Yuval Alaluf, Or Patashnik, and Daniel Cohen-Or. 2021. Only a Matter of Style: Age Transformation Using a Style-Based Regression Model.  
1075 <https://doi.org/10.48550/arXiv.2102.02754>
- 1076 [3] Yuval Alaluf, Or Patashnik, Zongze Wu, Asif Zamir, Eli Shechtman, Dani Lischinski, and Daniel Cohen-Or. 2022. Third Time's the Charm? Image  
1077 and Video Editing with StyleGAN3. <https://doi.org/10.48550/arXiv.2201.13433>
- 1078 [4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International  
1079 Conference on Machine Learning*. PMLR, 214–223.
- 1080 [5] A.h. Bermano, R. Gal, Y. Alaluf, R. Mokady, Y. Nitzan, O. Tov, O. Patashnik, and D. Cohen-Or. 2022. State-of-the-Art in the Architecture, Methods  
1081 and Applications of StyleGAN. *Computer Graphics Forum* 41, 2 (2022), 591–611. <https://doi.org/10.1111/cgf.14503>
- 1082 [6] Jianlong Chang, Gaofeng Meng, Lingfeng Wang, Shiming Xiang, and Chunhong Pan. 2020. Deep Self-Evolution Clustering. *IEEE Transactions on  
1083 Pattern Analysis and Machine Intelligence* 42, 4 (2020), 809–823. <https://doi.org/10.1109/TPAMI.2018.2889949>
- 1084 [7] Marie Charpentier, Elise Huchard, Anja Widdig, Olivier Gimenez, Bettina Sallé, Peter Kappeler, Julien Renoult, and L. Fusani. 2012. Distribution of  
1085 Affiliative Behavior Across Kin Classes and Their Fitness Consequences in Mandrills. *Ethology* 118 (Dec. 2012). <https://doi.org/10.1111/eth.12026>
- 1086 [8] Marie JE Charpentier, Clémence Poirotte, Berta Roura-Torres, Paul Amblard-Rambert, Eric Willaume, Peter M Kappeler, François Rousset, and  
1087 Julien P Renoult. 2022. Mandrill mothers associate with infants who look like their own offspring using phenotype matching. *eLife* 11 (Nov. 2022),  
1088 e79417. <https://doi.org/10.7554/eLife.79417>
- 1089 [9] M. J. E. Charpentier, M. Harté, C. Poirotte, J. Meric de Bellefon, B. Laubi, P. M. Kappeler, and J. P. Renoult. 2020. Same father, same face: Deep learning  
1090 reveals selection for signaling kinship in a wild primate. *Science Advances* 6, 22 (May 2020), eaba3274. <https://doi.org/10.1126/sciadv.aba3274>
- 1091 [10] Nicolas M. Dibot, Sonia Tiew, Tamra C. Mendelson, William Puech, and Julien P. Renoult. 2023. Sparsity in an artificial neural network predicts  
1092 beauty: Towards a model of processing-based aesthetics. *PLOS Computational Biology* 19, 12 (Dec. 2023), e1011703. <https://doi.org/10.1371/journal.pcbi.1011703>

- 1093 pcbi.1011703
- 1094 [11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative  
1095 adversarial nets. *Advances in neural information processing systems* 27 (2014).
- 1096 [12] Yseult Hejja-Brichard, Kara Million, Julien P. Renoult, and Tamra C. Mendelson. 2023. Using generative artificial intelligence to test hypotheses  
1097 about animal signal evolution: A case study in an ornamented fish. *bioRxiv* (2023), 2023–03.
- 1098 [13] W. C. Osman Hill. 1953. *Primates: comparative anatomy and taxonomy : a monograph / by W.C. Osman Hill*. At the University Press, Edinburgh.
- 1099 [14] Johannes Hirn, José Enrique García, Alicia Montesinos-Navarro, Ricardo Sánchez-Martin, Veronica Sanz, and Miguel Verdú. 2022. A deep  
1100 Generative Artificial Intelligence system to predict species coexistence patterns. *Methods in Ecology and Evolution* 13, 5 (2022), 1052–1061.  
1101 <https://doi.org/10.1111/2041-210X.13827>
- 1102 [15] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017.  
1103 MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. <https://doi.org/10.48550/arXiv.1704.04861>
- 1104 [16] Samuel V. Hulse, Julien P. Renoult, and Tamra C. Mendelson. 2022. Using deep neural networks to model similarity between visual patterns:  
1105 Application to fish sexual signals. *Ecological Informatics* 67 (2022), 101486.
- 1106 [17] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. GANSpace: Discovering Interpretable GAN Controls. <https://doi.org/10.48550/arXiv.2004.02546>
- 1107 [18] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation.  
1108 <https://doi.org/10.48550/arXiv.1710.10196>
- 1109 [19] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-Free Generative Adversarial  
1110 Networks. In *Proc. NeurIPS*.
- 1111 [20] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. <https://doi.org/10.48550/arXiv.1812.04948>
- 1112 [21] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of  
1113 StyleGAN. <https://doi.org/10.48550/arXiv.1912.04958>
- 1114 [22] Diederik P. Kingma and Max Welling. 2022. Auto-Encoding Variational Bayes. <https://doi.org/10.48550/arXiv.1312.6114>
- 1115 [23] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.
- 1116 [24] Qi Mao and Siwei Ma. 2023. Enhancing Style-Guided Image-to-Image Translation via Self-Supervised Metric Learning. *IEEE Transactions on  
1117 Multimedia* 25 (2023), 8511–8526. <https://doi.org/10.1109/TMM.2023.3238313>
- 1118 [25] Wendong Mao, Shuai Yang, Huihong Shi, Jiaying Liu, and Zhongfeng Wang. 2023. Intelligent Typography: Artistic Text Style Transfer for Complex  
1119 Texture and Structure. *IEEE Transactions on Multimedia* 25 (2023), 6485–6498. <https://doi.org/10.1109/TMM.2022.3209870>
- 1120 [26] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2022. Image Segmentation Using Deep  
1121 Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 7 (2022), 3523–3542. <https://doi.org/10.1109/TPAMI.2021.3059968>
- 1122 [27] Sarah Nila, Pierre-Andre Crochet, Julien Barthes, Puji Rianti, Berry Juliandi, Bambang Suryobroto, and Michel Raymond. 2019. Male Homosexual  
1123 Preference: Femininity and the Older Brother Effect in Indonesia. *Evolutionary Psychology* 17, 4 (Oct. 2019), 1474704919880701. <https://doi.org/10.1177/1474704919880701>
- 1124 [28] Yotam Nitzan, Rinon Gal, Ofir Brenner, and Daniel Cohen-Or. 2022. LARGE: Latent-Based Regression Through GAN Semantics. 19239–19249.
- 1125 [29] Julien P. Renoult, H. Martin Schaefer, Bettina Sallé, and Marie J. E. Charpentier. 2011. The Evolution of the Multicoloured Face of Mandrills: Insights  
1126 from the Perceptual Space of Colour Vision. *PLOS ONE* 6, 12 (Dec. 2011), e29117. <https://doi.org/10.1371/journal.pone.0029117>
- 1127 [30] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in Style: A StyleGAN  
1128 Encoder for Image-to-Image Translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2287–2296.
- 1129 [31] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in Style: a StyleGAN  
1130 Encoder for Image-to-Image Translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- 1131 [32] Amédée Roy, Ronan Fablet, and Sophie Lanco Bertrand. 2022. Using generative adversarial networks (GAN) to simulate central-place foraging  
1132 trajectories. *Methods in Ecology and Evolution* 13, 6 (2022), 1275–1287. <https://doi.org/10.1111/2041-210X.13853>
- 1133 [33] Yujun Shen, Jinjin Gu, Xiaou Tang, and Bolei Zhou. 2020. Interpreting the Latent Space of GANs for Semantic Face Editing.
- 1134 [34] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep Unsupervised Learning using Nonequilibrium  
1135 Thermodynamics. <https://doi.org/10.48550/arXiv.1503.03585>
- 1136 [35] Katharine Abernethy (University of Stirling) and Fiona Maisels (Wildlife Conservation Society). 2016. IUCN Red List of Threatened Species:  
1137 Mandrillus sphinx. *IUCN Red List of Threatened Species* (Nov. 2016).
- 1138 [36] Laszlo Talas, John G. Fennell, Karin Kjærsmo, Innes C. Cuthill, Nicholas E. Scott-Samuel, and Roland J. Baddeley. 2020. CamoGAN: Evolving optimum  
1139 camouflage with Generative Adversarial Networks. *Methods in Ecology and Evolution* 11, 2 (Feb. 2020), 240–247. <https://doi.org/10.1111/2041-210X.13334>
- 1140 [37] B. Tiddeman, M. Burt, and D. Perrett. 2001. Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and  
1141 Applications* 21, 5 (July 2001), 42–50. <https://doi.org/10.1109/38.946630>
- 1142 [38] Bernard Tiddeman, Neil Duffy, and Graham Rabej. 2001. A general method for overlap control in image warping. *Computers & Graphics* 25, 1 (Feb.  
1143 2001), 59–66. [https://doi.org/10.1016/S0097-8493\(00\)00107-2](https://doi.org/10.1016/S0097-8493(00)00107-2)
- 1144 Manuscript submitted to ACM

- 1145 [39] Sonia Tieo, Jules Dezeure, Anna Cryer, Pascal Lepou, Marie J.E. Charpentier, and Julien Renoult. 2023. Social and sexual consequences of facial  
1146 femininity in a non-human primate. *iScience* 26, 10 (Oct. 2023), 107901.
- 1147 [40] Sonia Tieo, Jules Dezeure, Anna Cryer, Pascal Lepou, Marie JE Charpentier, and Julien P. Renoult. 2023. Social and sexual consequences of facial  
1148 femininity in a non-human primate. *iScience* 26, 10 (2023).
- 1149 [41] Sonia Tieo, Claudia Ximena Restrepo-Ortiz, Berta Roura-Torres, Loic Sauvadet, Mélanie Harté, Marie J.E. Charpentier, and Julien P. Renoult. 2023.  
1150 The Mandrillus Face Database: A portrait image database for individual and sex recognition, and age prediction in a non-human primate. *Data in*  
1151 *Brief* 47 (2023), 108939. <https://doi.org/10.1016/j.dib.2023.108939>
- 1152 [42] Paul Tresson, Dominique Carval, Philippe Tixier, and William Puech. 2021. Hierarchical Classification of Very Small Objects: Application to the  
1153 Detection of Arthropod Species. *IEEE Access* 9 (2021), 63925–63932. <https://doi.org/10.1109/ACCESS.2021.3075293>
- 1154 [43] Cédric Villani. 2009. *Optimal Transport*. Grundlehren der mathematischen Wissenschaften, Vol. 338. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-540-71050-9>
- 1155 [44] Andrey Voynov and Artem Babenko. 2020. Unsupervised Discovery of Interpretable Directions in the GAN Latent Space. In *Proceedings of the 37th*  
1156 *International Conference on Machine Learning*. PMLR, 9786–9796.
- 1157 [45] Qingbo Wu, Hongliang Li, Zhou Wang, Fanman Meng, Bing Luo, Wei Li, and King N. Ngan. 2017. Blind Image Quality Assessment Based on  
1158 Rank-Order Regularized Regression. *IEEE Transactions on Multimedia* 19, 11 (2017), 2490–2504. <https://doi.org/10.1109/TMM.2017.2700206>
- 1159 [46] Zongze Wu, Dani Lischinski, and Eli Shechtman. 2020. StyleSpace Analysis: Disentangled Controls for StyleGAN Image Generation. <https://doi.org/10.48550/arXiv.2011.12799>
- 1160 [47] Yu Yu, Weibin Zhang, and Yun Deng. 2021. *Frechet Inception Distance (FID) for Evaluating GANs*.
- 1161 [48] Weiwei Zhang, Jian Sun, and Xiaoou Tang. 2011. From Tiger to Panda: Animal Head Detection. *IEEE Transactions on Image Processing* 20, 6 (2011),  
1162 1696–1708. <https://doi.org/10.1109/TIP.2010.2099126>
- 1163
- 1164
- 1165
- 1166
- 1167
- 1168
- 1169
- 1170
- 1171
- 1172
- 1173
- 1174
- 1175
- 1176
- 1177
- 1178
- 1179
- 1180
- 1181
- 1182
- 1183
- 1184
- 1185
- 1186
- 1187
- 1188
- 1189
- 1190
- 1191
- 1192
- 1193
- 1194
- 1195
- 1196

**CHAPITRE 2 : ATTRACTIVITY OF FACIAL  
FEMINITY OF A NON-HUMAN PRIMATE :  
EXPERIMENTAL EVIDENCE BASED ON  
GENERATIVE AI**



[128] : TIEO et al. (2023), « Social and sexual consequences of facial femininity in a non-human primate »

Le deuxième chapitre de cette thèse s'appuie sur le développement méthodologique réalisé dans le premier chapitre pour tester une hypothèse de préférence des mandrills mâles pour les femelles les moins féminines, basée sur des travaux récents [128].

Ainsi, nous avons montré à des mandrills mâles semi-captifs des paires d'images générées artificiellement de visages de mandrills femelles variant uniquement selon le paramètre de féminité conçu et décrit dans le chapitre précédent.

Après avoir mesuré le temps d'observation respectif pour chacune des images de la paire, nous avons validé l'hypothèse.



# Attractivity of facial femininity of a non-human primate : experimental evidence based on generative AI

Nicolas M. Dibot<sup>1,\*</sup>, Alice Baniel<sup>2</sup>, Marie J.E. Charpentier<sup>2</sup>, William  
Puech<sup>3</sup>, and Julien P. Renoult<sup>1</sup>

<sup>1</sup>CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>2</sup>Institut des Sciences de l'Évolution de Montpellier UMR5554, CNRS, IRD, EPHE, Université  
de Montpellier, Montpellier, France

<sup>3</sup>LIRMM, Univ. Montpellier, CNRS, Montpellier, France

\*Corresponding author: [nicolas.dibot@cefe.cnrs.fr](mailto:nicolas.dibot@cefe.cnrs.fr)

## Abstract

While in humans, the most feminine female faces are considered more attractive than the least feminine, recent research has shown that in mandrills, *Mandrillus sphinx*, a West African primate species, it's the opposite. However, previous evidence comes from the study of a wild population, based on correlative analyses, which prevents the exclusion of confounding factors. Here, we propose to leverage the power of generative artificial intelligence by experimentally testing the attractiveness of female mandrill faces to males. By generating images of synthetic faces, edited along a femininity axis using the StyleGAN3 generative model, we show that less feminine female mandrills are indeed preferred. This work provides the first experimental validation that male preference for femininity is not universal and can be reversed in some species, but also demonstrates the potential of recent developments in Generative AI to better understand animals' visual communication.

**Keywords:** mandrills, femininity, attractivity, socio-sexual behavior, mate choice, faces, generative artificial intelligence, StyleGAN3, editing, experimental approach.

## 1. Introduction

In human, facial femininity, the property to be representative of the “female” perceptual category (Fraccaro et al., 2010; Perrett et al., 1998; Rhodes et al., 2003) is a major determinant of sexual preferences. It is notably a key element of attractiveness, and thus is associated with a greater number of sexual partners (Rhodes et al., 2005). There are two possible explanations for the preference for feminine faces. Either femininity is a perceptual construct that makes it easy to identify traits beneficial to reproduction, or

it is a by-product of a general attraction to prototypes, the average representatives of a category. Indeed, beyond faces, people are attracted to all kinds of prototypes (animals, objects, etc.). As in humans, both hypotheses predict the same optimum, i.e. a preference by men for feminine women, identifying the relative contribution of the two underlying mechanisms has so far remained challenging , and supporters of each hypothesis continue to debate in the literature (Jokela, 2009; Little et al., 2014) .

However, a recent study on mandrill identified an inverse relationship between preference and femininity. Less feminine females were more approached and aggressed by individuals of both sexes, and received more copulation from males (Tieo, Dezeure et al., 2023). The preference for femininity would have been reversed, as in this species it would be beneficial for females to have masculine traits. This result would suggest that the preference for feminine faces is not universal, and therefore probably not a perceptual bias, but rather an adaptation shaped and shapable according to the ecology of the species.

The work that produced this result was nevertheless based on correlative studies carried out on a wild population. This type of approach makes it difficult to identify confounding factors and control them. Thus, it is not possible to isolate a single explanatory variable from the results. The correlative approach cannot therefore prove the causality of a relationship. To do so, we need to resort to experimentation, which requires manipulating the femininity of faces. To generate images of faces varying specifically in femininity, morphing techniques using traditional portraits (Bousquet & Kaminski, 2022; Nila et al., 2019; Rowland & Perrett, 1995) is a widespread solution in human studies. However, morphing is unable to capture all the variables that can influence femininity.

Thanks to the recent advances in generative artificial intelligence, it is now possible to create highly realistic synthetic stimuli (Karras et al., 2021), but also to manipulate them along a specific axis of variation that is decorrelated from the other variables (Bermano et al., 2022). One of the best models that can achieve such disentanglement between axes of variation is StyleGAN3 (Karras et al., 2021). GANs are a class of generative AI model where one neural network, called the generator, tries to synthesize the most realistic images possible, and another network, the discriminator, tries to differentiate real images from those of the generator. These two networks are optimized simultaneously in a kind of "arms race". StyleGAN is a type of GAN architecture that enables particularly realistic images to be produced, especially of faces, because a third neural network is connected to the generator and is specialized in detecting the style characteristics of images.

In this work, we carry out an experiment with captive male mandrills who are shown pairs of female face images varying only in degree of femininity, thanks to StyleGAN3. It is hypothesized that mandrills will look more at the less feminine images of the pairs, and that the greater the difference in femininity, the greater the difference in looking time.

## 2. Materials and methods

69

The aim of this study is to investigate the influence of femininity via choice tests between two portraits of the same female mandrill modified along a femininity axis. In section 2.1, we explain how we constructed these synthetic stimuli, then in 2.2 describe how we designed our experiment, and in 2.3 show how we analyzed the results.

70

71

72

73

### 2.1 Generation of stimuli

74

#### 2.1.1 Images of real faces

75

For study, we used the Mandrillus Face Database (MFD), a database of photographic portraits of mandrills created by the Mandrillus Project, a long-term monitoring project of a population of wild mandrills habituated to human presence in southeastern Gabon. MFD includes 29,395 portrait photographs of 397 different male and female mandrills (Tieo, Restrepo-Ortiz et al., 2023). Portrait photographs have been taken by field assistants using a camera and a long focal lens, and have been then manually processed to align the eyes horizontally and cropped to keep a square centered on the head. Each image is annotated with the date of shooting, the individual’s identity, its sex and date of birth, as well as two variables describing the image quality (score ranging from 0 to 3) and the orientation of the individual’s head (frontal or profile view). This database is in open access. For our work, we analyzed a subset of 18,358 photos, retaining only good-quality images (i.e. discarding image quality score 0) of adult individuals (i.e. aged 9 years and above) in frontal view.

76

77

78

79

80

81

82

83

84

85

86

87

88

#### 2.1.2 Images of synthetic faces

89

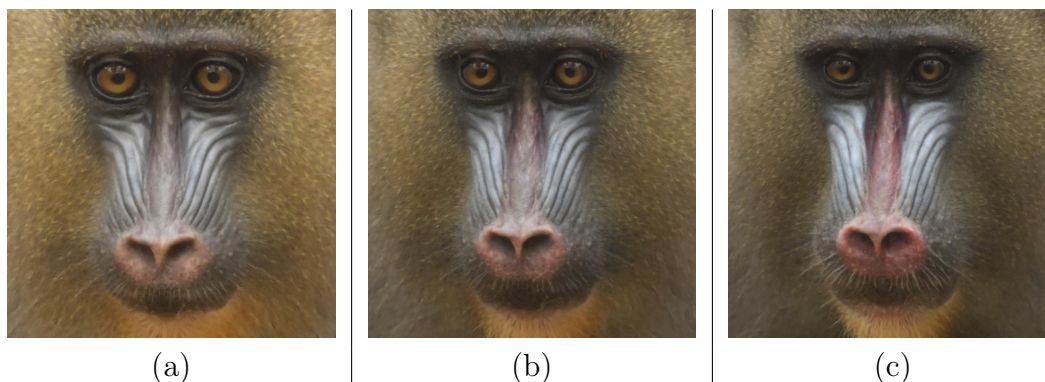


Figure 1: example editing of an image with our method, more feminine on the left (a), unedited on the middle (b) less feminine (i.e more masculine) on the right (c)

The details of the method is described in a dedicated methodological article (Dibot et al., in prep). Briefly, a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) was first trained to produce synthetic images of mandrill faces from the database mentioned in the previous section. We used a GAN architecture, StyleGAN3, which is

90

91

92

93

particularly effective for creating realistic synthetic images , but also for creating latent feature spaces that linearize descriptive variables of interest in the images. The network was trained using the weights of a model pre-trained on human faces. The trained model was used to generate 100,000 synthetic faces, that is, realistic faces of mandrills that do not exist. Each synthetic image is associated with a latent vector, i.e. its coordinates in the latent space of the GAN.

Next, we trained a convolutional neural network (CNN) to classify images of the MFD database according to the sex and used this CNN to classify the synthetic images according to their sex. We then trained a Support Vector Machine (SVM) to separate the images of males and females in the GAN latent space. We then calculated the sex editing vector as a vector perpendicular to the support vectors. Thus, by translating an image along this vector in the latent space, we were able to increase or decrease the femininity of the depicted face, independently of other variables: we call this manipulation editing.

Finally, in order to control the magnitude of the editing in the latent space, we projected all the images of real faces available in MFD onto the sex editing vector. To do so, we trained an encoder (Richardson et al., 2021) specifically designed to project real images into the StyleGAN3's latent space. Knowing the distribution of real faces along the sex editing vector, we were able to edit synthetic images along a range of femininity values that are biologically realistic, empirically determined to +2 or -2 standard deviations around the mean, for males and females respectively.

The femininity values obtained on this axis are called "sex levels", and are to be interpreted relatively to each other, or with regard to the distribution of real images, since their score in absolute terms has no particular meaning. The higher the sex level, the more masculine the face, the lower the sex level, the more feminine the face (Cf. Fig. 1).

### 2.1.3 Validation by experts

To validate our editing of femininity, we asked 10 experts to carry out a pairwise choice test. The experts were Gabonese or French field assistants or researchers working for the Mandrillus Project, selected for their ability to recognize mandrills individually. Each expert was asked to select the most feminine face in each of 100 pairs of portrait images. To avoid bias in relation to the experts' perception of the term "femininity", the exact question was "Click on the mandrill that looks most like a female" (original sentence in French: "Cliquez sur le mandrill qui 'fait le plus femelle'). The images of a given pair displayed the same face varying only in its sex level. Pairs were created by randomly sampling images from a set of 31 female images, created from 5 real images. edited by increasing and decreasing their sex level.

We then calculated a score for each of the 31 images, using the Elo score method (Goodspeed, 2017), originally designed to rank chess players according to their won or lost games and the level of their opponents. Here, we consider each image as a player and

each pair presented to the same expert as a match. 133

Finally, we calculated the Spearman correlation coefficient between sex levels (which 134  
do not follow a normal distribution) and Elo scores, to assess the extent to which editing 135  
for sex with our method was correlated with experts' perception of femininity. 136

## 2.2 Looking time experiments 137

### 2.2.1 Experimental population 138

We carried the experiment with the semi-captive mandrills of the Centre International de 139  
Recherches Médicales de Franceville (CIRMF) in Gabon. The advantage of working with 140  
these animals is firstly to have an environment conducive to experiments under controlled 141  
conditions, and secondly to avoid a familiarity bias between the stimuli and the wild 142  
mandrill population of the Mandrillus Project. 143

These individuals come from a population captured in the wild in the 80s, and are 144  
divided into 3 independent enclosures (enclosures 1, 2 and 3). An enclosure consists 145  
of a fragment of rainforest, spanning several hectares, where around sixty individuals 146  
live. The animals are fed daily in cages on the periphery of the enclosures, so that 147  
they are accustomed to entering these enclosed spaces where the experiments are carried 148  
out. However, outside the feeding period, they do not enter the cages and remain in the 149  
forest. Each cage is made up of two parts: one where the individuals can enter from the 150  
enclosure, the feeding room, and a second where they can only enter from the first part, 151  
the experiment room. The openings between the cages and the enclosure can be managed 152  
from the outside. Enclosure 1 has a single cage. Enclosures 2 and 3 have a common cage 153  
at the boundary between the two enclosures, but Enclosure 2 also has a second cage of 154  
its own (see a plan in S3). 155

In our work, we are specifically interested in adult male mandrills. We carried out 156  
these experiments with 23 individuals: 11 in enclosure 1, 11 in enclosure 2 and 1 in 157  
enclosure 3. The ages of the captured mandrills ranged from 9 to 22 years (mean: 13.7 158  
years), as mandrills are considered adults between 9 and 10 years of age (Setchell et al., 159  
2005). Some individuals were tested several times. 160

### 2.2.2 Experimental design 161

The experiments are carried out at feeding time in the morning. Individuals enter to feed 162  
in the first part of the cages, the feeding room, where a large amount of fruit has been 163  
placed. One individual is isolated in the second part, the experiment room, to carry out 164  
the experiment. A video describes this in S1. Four bananas are scattered on the floor to 165  
attract the individual and encourage him to roam the space. 166

Two Samsung U32J590UQP - UJ59 Series 4K 32" screens are positioned on one side 167  
of the cage in wooden and Plexiglas boxes (Cf. Fig. 2) to display a pair of stimuli of the 168  
same size female image edited with two different femininity scores. As the screens are 169

larger than the images, a gray background is positioned behind the images. The shade of gray avoids the mirror effect that would be caused by a black color. Two cameras were positioned to record the tests. A Sony Handycam AX53 4K camera filmed the individuals from the front (positioned between the two screens), and a GoPro HERO11 camera filmed the individuals at 45° above the screens.

When an individual is isolated in the experiment room, we start a 20-minute session. This session begins with 3 minutes of totally gray screen, to allow the individual to calm down from the stress of being isolated. A first trial then begins, showing a pair of images for 5 minutes. This is followed by 1 minute of gray screen, another 5 minutes of a new pair, then one minute of gray screen, a third pair for 5 minutes and then the end of the trial (Cf. Fig. 3). For each pair, the right or left screen projecting the more feminine face is chosen randomly.



Figure 2: **An experimental session** A mandrill is in the cage with two screens showing the same female, more masculine on the left and more feminine on the right.

In all, we carried out 46 sessions totalling 138 trials. From these, we excluded 28 trials during which the tested mandrill did not look at either of the screens, or only looked at one of the two screens.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	...
grey screen			first pair						second pair						third pair				end		

Figure 3: **chronological sequence of a session** the numbers on the first line correspond to the minutes of the session.

### 2.2.3 Statistical analysis of experiments 185

We run a binomial general linear mixed model (Binomial GLMM) to test our hypotheses. 186  
The response variable is a success/failure binomial, where success is the number of seconds 187  
spent looking at the least feminine screen, and failure is the number of seconds spent 188  
looking at the most feminine screen. 189

The variables tested are those that have a significant effect on the amount of time 190  
mandrills spend looking at the screens, in addition to the sex level of the less feminine 191  
image and the difference in sex level between the two images, which correspond directly 192  
to our hypotheses. The individual effect is taken into account as a random effect. The 193  
age status is "sub-adult" if the individual is between 9 and 10 years old, and "adult" if it 194  
is over 10 years old. We also test the interaction effect between the sex level of the least 195  
feminine image and the difference in sex level between the two images . 196

## 3. Results 197

### 3.1 Expert evaluation of our sex-level metric 198

We first evaluated the proposed editing method with mandrill experts. 199

The Spearman correlation coefficient between sex levels (the computed femininity) and 200  
ELO scores (the empirically estimated femininity) was -0.56 ( $P=0.001$ )(Cf. Fig. 4). 201

A high sex level indicates a less feminine image, while a high ELO score indicates a 202  
more feminine image, thereby explaining the negative relationship. 203

### 3.2 Descriptive statistics and explanatory analysis 204

Mandrills spent between 2 and 157 seconds looking at both screens during a trial, with 205  
an average of 18.5 seconds. With a trial lasting 350 seconds, they spent on average 5.3% 206  
of the time looking at the screens. 207

Concerning the individual effect, we performed a Dunn's post-hoc test which identified 208  
that one individual, #33C, looked significantly less at the screens than 3 other individuals: 209  
#10F5H2, #12 and #12A5E ( $p < 0.05$  in all 3 cases). There is no difference between the 210  
observation times of all the other mandrills in pairs. Concerning age status, we specify 211  
that we studied this in addition to age in year, as we had missing data for age in year 212  
for 3 individuals. Regarding the cage effect, we performed a Dunn's post-hoc test, which 213  
identified that individuals in the cage of enclosure 1 looked significantly more at the 214  
screens than those in the cage of enclosure 2, which is not shared with enclosure 3 (Cf. 215  
Fig. 1). 216

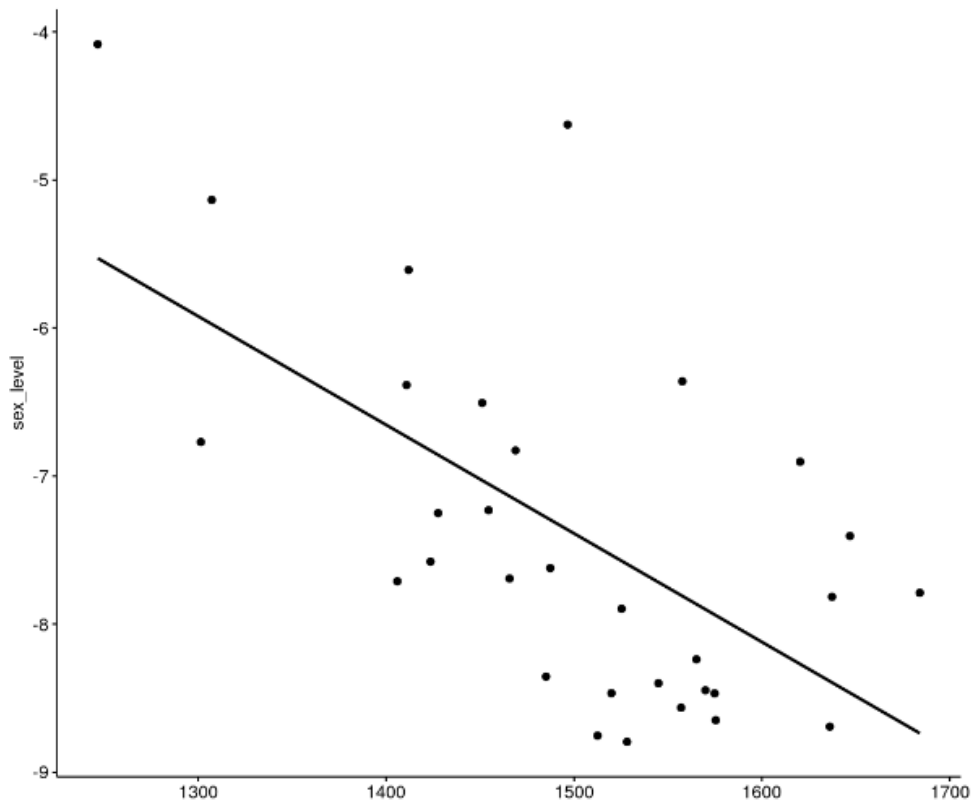


Figure 4: **Scatter plot between Elo rating and sex level** The line represents the linear trend fitted using a linear regression model. The Spearman correlation coefficient is  $\rho = -0.56$  ( $p < 0.05$ )

### 3.3 Binomial model

217

The results of the binomial model, after removing the interaction between the sex level of the least feminine image and the difference in sex level between the two images, which was not significant in a preliminary model, show that mandrills look more at screens showing a less feminine mandrill female than at screens showing a more feminine mandrill female (Cf. Fig. 5). In addition, the magnitude of the difference in sex level between the two images also influences gaze time for more feminine faces. Finally, when the experiment takes place in certain cages, the effect is also amplified. In contrast, the age status of the individual had no influence Cf. Tab. 2).

218  
219  
220  
221  
222  
223  
224  
225

## 4. Discussion

226

We tested the hypothesis of male mandrills' attractiveness to less feminine females, with a pairwise choice test. Males looked longer at the more feminine female faces in the pairs, in line with our prediction. Beyond this result, we used an innovative approach based on stimuli created with generative artificial intelligence. The results obtained demonstrate the relevance of this approach, reinforcing results obtained with a population of wild mandrills in previous work, but overcoming the limitations of this work. We also found an effect of the intensity of the difference in femininity between the two images of the

227  
228  
229  
230  
231  
232  
233



Variable	Test (non-normal distributions)	p
sex level of the least feminine screen	Spearman	0.09
laterality of the least feminine screen	Wilcoxon-Mann-Whitney	0.32
sex level difference between the two images	Spearman	0.20
original image of the pair	Kruskall-Wallis	0.24
order of trial in the session	Kruskall-Wallis	0.45
Individual identifier	Kruskall-Wallis	< <b>0.0005</b>
Individual age in years	Spearman	0.26
individual age status	Wilcoxon-Mann-Whitney	< <b>0.005</b>
individual rank	Kruskall-Wallis	0.25
number of sessions already completed by the individual	Kruskall-Wallis	0.17
cage where the test is carried out	Kruskall-Wallis	< <b>0.01</b>

Table 1: **Statistical tests carried out (second column) between the descriptive variables of the experiments (first column) and the time spent by individuals looking at the two screen images.** All tests take into account that the variables studied follow a non-normal distribution, previously tested with a Shapiro-Wilk test. A p-value is indicated in red if significant and in black otherwise (third column).

	Estimate	Std. Error	z value	Pr (> z )	significance
(Intercept)	0.13067	0.45669	0.286	0.77478	/
statut age (adult)	-0.30962	0.22058	-1.404	0.16042	/
cage (1)	0.78934	0.25151	3.138	0.00170	**
cage (2) status age (adult)	0.95832	0.17294	5.541	3e-08	***
sex level difference between the two images	0.14357	0.04813	2.983	0.00285	**
sex level of the less feminine screen	0.10674	0.04463	2.392	0.01677	*

Table 2: Results of the binomial general linear mixed model examining the probability that a mandrill looks toward the less feminine face image. The modality tested is indicated in brackets for categorical variables. Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

pair.. Furthermore, differences between different cages of the experimental set-up seem 234  
to have an effect on the preference for less feminine females. 235

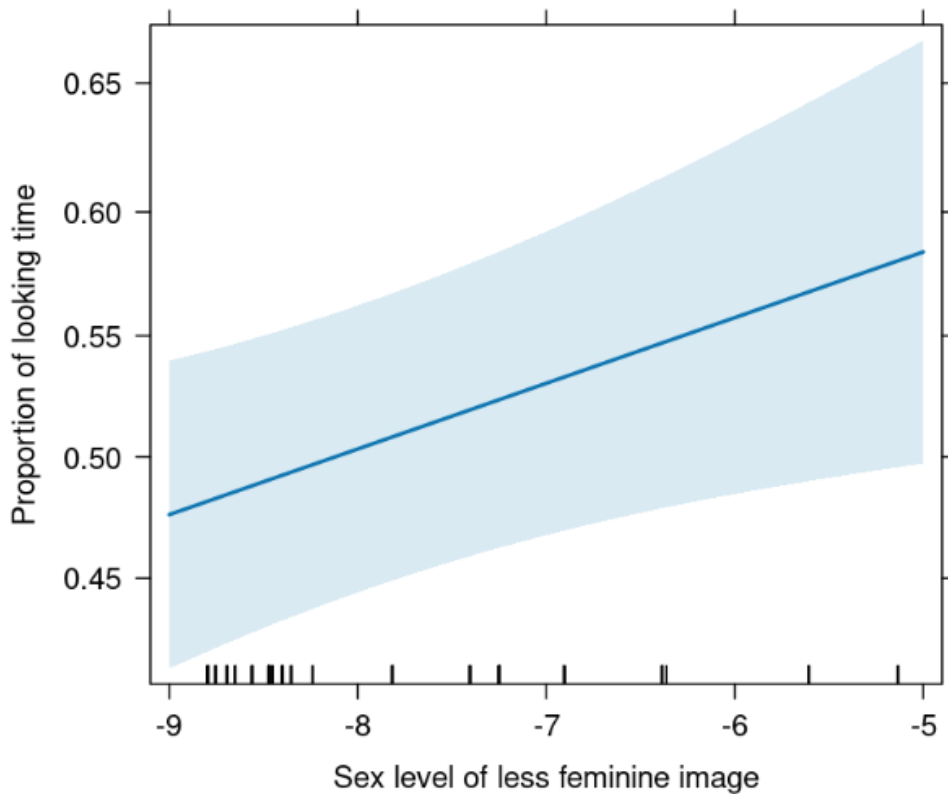


Figure 5: **Facial femininity score plotted against less feminine image looking time**  
 Blue line represents the line of best fit, and the shaded region represents 95% confidence intervals for this line.

#### 4.1 Preference for less feminine females

236

We have thus seen that mandrills look longer at less feminine female mandrills than at 237  
 more feminine female mandrills. This result suggests a preference in this direction. This 238  
 preference might be explained by the fact that in mandrills, a less feminine female face 239  
 is perceived as an indicator of quality for males. This result is in line with other work 240  
 in mandrills (Tieo, Dezeure et al., 2023) but is the opposite to what is observed in other 241  
 primates, notably in humans, where males seem to prefer more feminine female faces in 242  
 most contexts. Nevertheless, in humans, this preference is not universal: more feminine 243  
 women are, for example, considered less trustworthy (Little et al., 2014). Thus, this 244  
 valence switch seems very much linked to the social context in which the preference is 245  
 expressed. Masculine traits could thus be advantageous for female mandrills, as they 246  
 would indicate a greater ability to protect children or appropriate resources in a conflict 247  
 context. 248

These results suggest that in the case of preferences according to femininity level, 249  
 an explanation based on a perceptual bias linked to a preference for prototypes is not 250  
 likely. While prototypes, the average representatives of a category, have been argued to 251  
 be preferred because they are easy to form a mental representation (Winkielman et al., 252  
 2006)of, this does not seem to be the case for female prototypes, i.e. the most feminine 253  
 mandrills. Nevertheless, one possible area to explore that could challenge this is the 254

interaction between preference valence and fluency theory. Indeed, fluency could reinforce 255  
negative valence, put another way, the easier a prototype would be to process, the more 256  
negatively it would be judged (Ingendahl et al., n.d.). Thus, femininity would be judged 257  
negatively and thus favor masculine traits in females. 258

## 4.2 Generative AI and experimental design 259

Our work has also made it possible to evaluate the value of an approach based on modifying 260  
the femininity of faces using generative AI. The sex level derived from this modification is 261  
valid, as it is confirmed by a correlation with a score derived from expert evaluation. The 262  
correlation remains moderate (-0.56) but is significant and therefore sufficient to conclude 263  
on the relevance of our evaluation. Experts are trained to recognize individual mandrills, 264  
not to assess their sex. What's more, they are used to seeing mandrills in the wild, where 265  
the major difference between males and females is size, a particularly important difference 266  
for this species. Males are 2 to 3 times larger and heavier than females (Setchell & Jean 267  
Wickings, 2006). 268

We expected the magnitude of the difference between the two pair images to correlate 269  
with longer observation time for the less feminine females. Our results show no such 270  
correlation, which can be explained by two factors. Firstly, males may be able to differ- 271  
entiate the level of femininity no matter how small, so even a slight difference is obvious. 272  
Secondly, our experimental design may not be sufficiently adapted to test this. We could 273  
imagine modifying it by testing two types of difference between pairs: one where the sex 274  
levels are very different, another where the difference is very small. 275

The looking time paradigm on which our experimental approach is based has its lim- 276  
itations (Winters et al., 2015)). The main one is that looking time indicates choice rather 277  
than preference. These two concepts are not necessarily linked. For example, in another 278  
fictitious context, a mandrill may have a preference for bananas rather than mangoes. 279  
However, eating a banana could increase the risk of being attacked by another mandrill 280  
who would steal his banana by beating him. So, faced with a banana and one, this 281  
mandrill might choose the mango, even though it prefers bananas. 282

Finally, the results of the model show that the experiment room where the experiment 283  
takes place, among 3 different rooms, has an importance. This can be explained by the 284  
slightly different structure of these cages, impossible to control due to the constraints of 285  
the research center where we carried out the experiments. Indeed, the cages were not 286  
quite symmetrical, since on one side there was the enclosure with other mandrills that 287  
could distract the test individual, and on the other, there was the empty feeding room 288  
(Cf. S3). Although we randomized the side where the most feminine screen was located 289  
in each trial, this may not have been sufficient. Nevertheless, we added this cage variable 290  
to the model, allowing us to control it. 291

## References

- Bermano, A., Gal, R., Alaluf, Y., Mokady, R., Nitzan, Y., Tov, O., Patashnik, O., & Cohen-Or, D. (2022). State-of-the-Art in the Architecture, Methods and Applications of StyleGAN [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14503>]. *Computer Graphics Forum*, 41(2), 591–611. <https://doi.org/10.1111/cgf.14503>
- Bousquet, C. A. H., & Kaminski, G. (2022). Transforming faces to mimic natural kin: A comparison of different paradigms. *Behavior Research Methods*, 54(1), 13–25. <https://doi.org/10.3758/s13428-021-01614-5>
- Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated Male Preferences for Femininity in Female Faces and Voices [Publisher: SAGE Publications Inc]. *Evolutionary Psychology*, 8(3), 447–461. <https://doi.org/10.1177/147470491000800311>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June). Generative Adversarial Networks [arXiv:1406.2661 [cs, stat]]. <https://doi.org/10.48550/arXiv.1406.2661>
- Goodspeed, R. (2017). Research note: An evaluation of the Elo algorithm for pairwise visual assessment surveys. *Landscape and Urban Planning*, 157, 131–137. <https://doi.org/10.1016/j.landurbplan.2016.06.009>
- Ingendahl, M., Propheter, N., & Vogel, T. (n.d.). The role of category valence in prototype preference [Publisher: Routledge eprint: <https://doi.org/10.1080/02699931.2024.2335536>]. *Cognition and Emotion*, 0(0), 1–7. <https://doi.org/10.1080/02699931.2024.2335536>
- Jokela, M. (2009). Physical attractiveness and reproductive success in humans: Evidence from the late 20th century United States. *Evolution and Human Behavior*, 30(5), 342–350. <https://doi.org/10.1016/j.evolhumbehav.2009.03.006>
- Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021, October). Alias-Free Generative Adversarial Networks [arXiv:2106.12423 [cs, stat]]. <https://doi.org/10.48550/arXiv.2106.12423>
- Little, A. C., Jones, B. C., Feinberg, D. R., & Perrett, D. I. (2014). Men’s strategic preferences for femininity in female faces [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12043>]. *British Journal of Psychology*, 105(3), 364–381. <https://doi.org/10.1111/bjop.12043>
- Nila, S., Crochet, P.-A., Barthes, J., Rianti, P., Juliandi, B., Suryobroto, B., & Raymond, M. (2019). Male Homosexual Preference: Femininity and the Older Brother Effect in Indonesia. *Evolutionary Psychology: An International Journal of Evolutionary Approaches to Psychology and Behavior*, 17(4), 1474704919880701. <https://doi.org/10.1177/1474704919880701>
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L., & Akamatsu, S. (1998). Effects of sexual dimorphism

- on facial attractiveness [Publisher: Nature Publishing Group]. *Nature*, 394(6696), 330  
884–887. <https://doi.org/10.1038/29772> 331
- Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism 332  
in human faces signal health? [Publisher: Royal Society]. *Proceedings of the Royal* 333  
*Society of London. Series B: Biological Sciences*, 270(suppl\_1), S93–S95. <https://doi.org/10.1098/rsbl.2003.0023> 334  
335
- Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual beha- 336  
vior: Does attractiveness enhance mating success? *Evolution and Human Behavior*, 337  
26(2), 186–201. <https://doi.org/10.1016/j.evolhumbehav.2004.08.014> 338
- Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., & Cohen- 339  
Or, D. (2021, April). Encoding in Style: A StyleGAN Encoder for Image-to-Image 340  
Translation [arXiv:2008.00951 [cs]]. <https://doi.org/10.48550/arXiv.2008.00951> 341  
Comment: Accepted to CVPR 2021, project page available at <https://eladrich.github.io/pixel2sty>
- Rowland, D., & Perrett, D. (1995). Manipulating Facial Appearance Through Shape and 343  
Color. *Computer Graphics and Applications, IEEE*, 15, 70–76. <https://doi.org/10.1109/38.403830> 344  
345
- Setchell, J. M., Charpentier, M., & Wickings, E. J. (2005). Sexual Selection and Repro- 346  
ductive Careers in Mandrills (*Mandrillus sphinx*) [Publisher: Springer]. *Behavioral* 347  
*Ecology and Sociobiology*, 58(5), 474–485. Retrieved July 17, 2024, from <https://www.jstor.org/stable/25063642> 348  
349
- Setchell, J. M., & Jean Wickings, E. (2006). Mate Choice in Male Mandrills (*Man-* 350  
*drillus sphinx*) [eprint: [https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439- 351  
\*0310.2006.01128.x\*\]. \*Ethology\*, 112\(1\), 91–99. \[https://doi.org/10.1111/j.1439- 352  
\\*0310.2006.01128.x\\* 353\]\(https://doi.org/10.1111/j.1439-0310.2006.01128.x\)](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.2006.01128.x)
- Tieo, S., Dezeure, J., Cryer, A., Lepou, P., Charpentier, M. J. E., & Renoult, J. P. 354  
(2023). Social and sexual consequences of facial femininity in a non-human primate. 355  
*iScience*, 26(10), 107901. <https://doi.org/10.1016/j.isci.2023.107901> 356
- Tieo, S., Restrepo-Ortiz, C. X., Roura-Torres, B., Sauvadet, L., Harté, M., Charpentier, 357  
M. J. E., & Renoult, J. P. (2023). The *Mandrillus* Face Database: A portrait image 358  
database for individual and sex recognition, and age prediction in a non-human 359  
primate. *Data in Brief*, 47, 108939. <https://doi.org/10.1016/j.dib.2023.108939> 360
- Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes Are 361  
Attractive Because They Are Easy on the Mind [Publisher: SAGE Publications 362  
Inc]. *Psychological Science*, 17(9), 799–806. [https://doi.org/10.1111/j.1467- 363  
\*9280.2006.01785.x\* 364](https://doi.org/10.1111/j.1467-9280.2006.01785.x)
- Winters, S., Dubuc, C., & Higham, J. P. (2015). Perspectives: The Looking Time Experi- 365  
mental Paradigm in Studies of Animal Visual Perception and Cognition [eprint: 366  
<https://onlinelibrary.wiley.com/doi/pdf/10.1111/eth.12378>]. *Ethology*, 121(7), 625– 367  
640. <https://doi.org/10.1111/eth.12378> 368

## 5. Supplementary material

369

<https://youtu.be/XyPEAAEYxQI>

370

**S1:** video of the experimental apparatus (explanation in french,)

371

372

<https://youtu.be/XnCiAxqvbkw>

373

**S2:** Camera shots from a session of 20mn, 3 successive trials

374

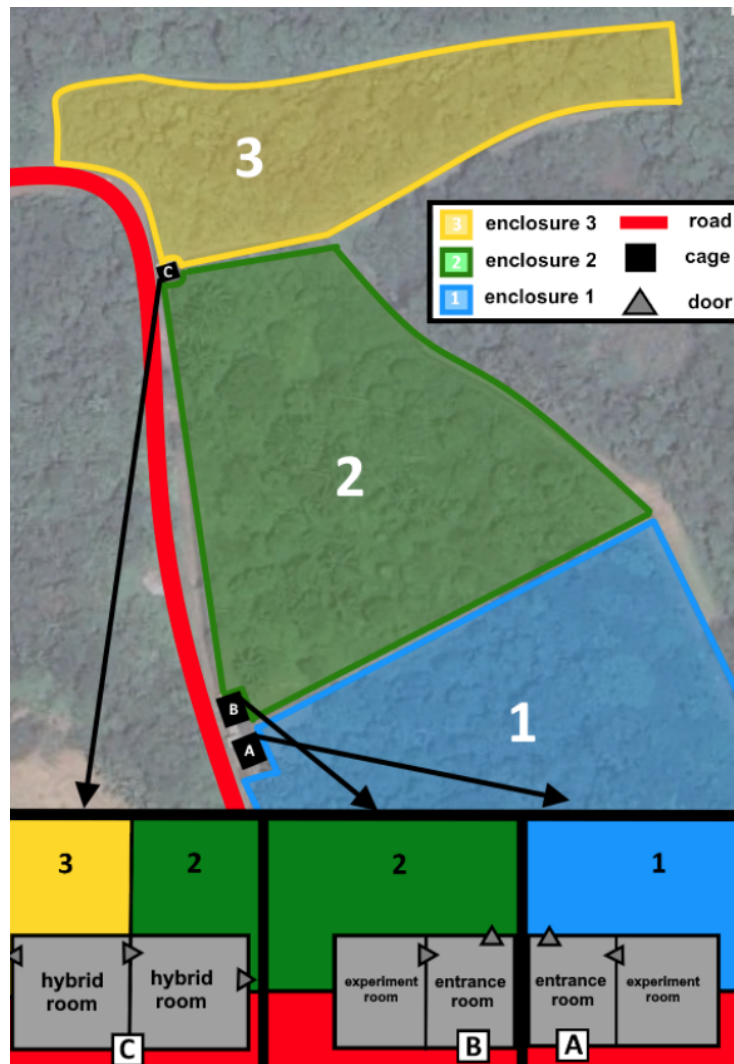


Figure 6: **S3:** enclosure plan in the top, cage plan in the bottom



**CHAPITRE 3 : SPARSITY IN AN  
ARTIFICIAL NEURAL NETWORK  
PREDICTS BEAUTY : TOWARDS A MODEL  
OF PROCESSING-BASED AESTHETICS**





Dans ce troisième chapitre, nous prenons du recul sur la capacité de l'intelligence artificielle à mettre en évidence des préférences, en nous intéressant à un paramètre d'attractivité autre que la féminité bien que pouvant avoir des liens avec celle-ci : la beauté.

Nous avons pu voir dans l'introduction que l'IA pouvait selon certains paramètres, servir de modèle de perception. Ici, nous utilisons un réseau de neurones convolutif dans cette perspective.

Notre hypothèse est que la fluence neuronale est liée à la perception de la beauté. Cette hypothèse est validée pour des sujets humains, en s'appuyant sur des images de visages humains, mais aussi d'oeuvres d'art, auxquelles sont associés des scores de beauté.

## RESEARCH ARTICLE

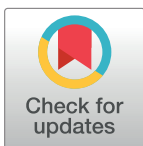
# Sparsity in an artificial neural network predicts beauty: Towards a model of processing-based aesthetics

Nicolas M. Dibot<sup>1,2\*</sup>, Sonia Tio<sup>1</sup>, Tamra C. Mendelson<sup>3</sup>, William Puech<sup>2</sup>, Julien P. Renault<sup>1</sup>

**1** CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Montpellier, France, **2** LIRMM, Univ. Montpellier, CNRS, Montpellier, France, **3** Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, Maryland, United States of America

☯ These authors contributed equally to this work.

\* [nicolas.dibot@cefe.cnrs.fr](mailto:nicolas.dibot@cefe.cnrs.fr)



## OPEN ACCESS

**Citation:** Dibot NM, Tio S, Mendelson TC, Puech W, Renault JP (2023) Sparsity in an artificial neural network predicts beauty: Towards a model of processing-based aesthetics. *PLoS Comput Biol* 19(12): e1011703. <https://doi.org/10.1371/journal.pcbi.1011703>

**Editor:** Roland W. Fleming, University of Giessen, GERMANY

**Received:** May 23, 2023

**Accepted:** November 20, 2023

**Published:** December 4, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1011703>

**Copyright:** © 2023 Dibot et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The code that allowed to obtain the results of this work is available on this Github repository: [https://github.com/NicolasDibot/sparsity\\_beauty](https://github.com/NicolasDibot/sparsity_beauty).

## Abstract

Generations of scientists have pursued the goal of defining beauty. While early scientists initially focused on objective criteria of beauty ('feature-based aesthetics'), philosophers and artists alike have since proposed that beauty arises from the interaction between the object and the individual who perceives it. The aesthetic theory of fluency formalizes this idea of interaction by proposing that beauty is determined by the efficiency of information processing in the perceiver's brain ('processing-based aesthetics'), and that efficient processing induces a positive aesthetic experience. The theory is supported by numerous psychological results, however, to date there is no quantitative predictive model to test it on a large scale. In this work, we propose to leverage the capacity of deep convolutional neural networks (DCNN) to model the processing of information in the brain by studying the link between beauty and neuronal sparsity, a measure of information processing efficiency. Whether analyzing pictures of faces, figurative or abstract art paintings, neuronal sparsity explains up to 28% of variance in beauty scores, and up to 47% when combined with a feature-based metric. However, we also found that sparsity is either positively or negatively correlated with beauty across the multiple layers of the DCNN. Our quantitative model stresses the importance of considering how information is processed, in addition to the content of that information, when predicting beauty, but also suggests an unexpectedly complex relationship between fluency and beauty.

## Author summary

Developing good predictive models of beauty requires understanding what happens in the brain when we find a person or an artwork beautiful. Recent theories in psychology emphasize the importance of considering how the brain processes features, in addition to the features themselves. Features that are efficiently processed by the brain, such as symmetry, fractality, or naturalness are generally perceived as visually attractive. In this study,

**Funding:** This study was funded by the Agence Nationale de la Recherche (ANR-20-CE02-0005-01) received by JPR and WP, by the National Science Foundation (NSF IOS 2026334) received by TM and JPR, and by the CNRS through the MITI interdisciplinary programs (Programme Interne Blanc MITI 2023.1 - Projet: DEEPCOM-L'intelligence artificielle pour étudier la communication) received by WP. ND and ST received a salary from ANR-20-CE02-0005-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

we leveraged the capacity of artificial intelligence to model information processing in the human brain, to evaluate how the beauty of human faces and artistic paintings can be predicted from the efficiency of the neural code. Our results show that the efficiency of information processing can explain approximately one-third of the perception of beauty and emphasize the importance of considering how information is processed when investigating beauty. Additionally, our use of artificial intelligence demonstrates the potential of this technology to help better understand complex human behaviors.

## Introduction

Understanding and predicting beauty has been a goal of humans for thousands of years. Early studies of beauty aimed to identify characteristics that make objects universally beautiful. Examples of such objective characteristics of beauty include the golden ratio, symmetry, or more complex measures such as fractality [1,2]. Later on, artists and philosophers focused on the subjective aspect of beauty, motivated by the recognition that the taste for the beautiful varies between individuals and even with age within individuals [3,4]. Then, over the XX<sup>th</sup> century, an interaction paradigm gained popularity that accounts for the fact that beauty is simultaneously universal and subjective [5]. Beauty is neither a property of objects nor an idiosyncrasy of the observer, rather it emerges from the interaction between the two [6]. It thus has been argued that the key to understanding beauty hides in the brain mechanisms underlying this interaction.

In psychology, the fluency theory of aesthetics [5] proposes that beauty is determined by how the brain processes information, and that fluency—a subjective sensation of ease in information processing—triggers a pleasurable aesthetic experience [7]. Fluency theory fits the interactionist paradigm very well because information processing is influenced by properties of both the objects and the sensory and cognitive systems that process them [8]. Fluency theory also emphasizes that describing information processing *per se*, in addition to the information being processed, is important to understanding beauty. Overall, the theory can explain remarkably well many, if not most, results of psychology on the phenomenon of beauty. For example, it can account for universal preferences for symmetrical, rounded, highly contrasting, and repeated forms (both through time and space), all of these being fluently processed [5], but it also explains context-dependent pleasure received from an acceleration of information flow (e.g., the ‘aha effect’ following suspense or in optical illusions [9]).

Fluency has been quantified in various ways, each with advantages but also limitations [10]. Reaction time, for example, is easy to measure but it is only indirectly linked to fluency, being constrained by motor mechanisms. Self-report of subjective ease more closely matches fluency, but it is generally poorly reliable [11]. Characteristics like symmetry and fractality have been often used as metrics of fluency [12]. However, using such characteristics re-embeds fluency into the strictly objectivist paradigm of beauty that dismisses the importance of the perceiver. Modeling beauty using the fluency theory thus requires developing metrics that describe the state of the sensory system or the brain while it is processing information. We qualify such metrics as “processing-based”. In contrast, “feature-based” metrics are focused on features, which are quantitative characteristics of objects, measured either directly from the object or as they are perceived. Contrary to processing-based metrics, which are interactive by construction, feature-based metrics can thus be either interactive or objective.

Several authors have argued that fluency could be modeled through the concept of efficiency, which describes the capacity to perform a task with an optimal use of neuronal

resources [8,13]. In neuroscience, efficiency is often measured through the sparsity of the neuronal code. A sparse code is one in which the vast majority of neurons are at rest, and only a few, highly specialized neurons are strongly activated [14]. Using the sparsity of neuronal activations to estimate fluency fits the prediction of Winkielman [15], that “fluent patterns should be represented by more extreme values of activation”. Accordingly, a previous study found that sparsity of neuronal activations in a model of the primary visual cortex was positively correlated with the attractiveness of the female face as rated by men [16]. Using a similar approach, Holzeleitner et al. [17] found that sparsity was the best predictor of face attractiveness as compared to body mass index, sexual dimorphism, averageness and asymmetry. Sparsity is also negatively correlated with aversiveness. Images of abstract patterns with a lower degree of sparsity are more highly aversive to human subjects [18]. While these previous studies collectively support a link between attractiveness and activation sparsity in the primary visual cortex, to what extent beauty and attractiveness can be explained by activation sparsity as measured in the visual system including, but not limited to the primary visual cortex has so far remained unexplored.

We propose to leverage the capacity of artificial neural networks to model the human visual system to investigate the link between neuronal activation sparsity. Some artificial intelligence architectures, and in particular Deep Convolutional Neural Networks (DCNN) can fulfill this role. DCNNs and the visual cortex have similarities in how they process information [19,20]. Accordingly, patterns of neuronal activations within a DCNN have been shown to predict those recorded in the visual cortex of humans [20–24]. Furthermore, previous studies have shown that metrics summarizing the processing of visual information by DCNNs correlate with self-reported assessment of this processing; for example, the mean activation per layer predicts the perceived complexity of an image [25].

Previous studies in aesthetic science have used DCNNs (for reviews, see [26,27]), for example to recognize artworks and classify artistic paintings by their style (e.g., [28]). DCNNs are also able to predict mean (e.g., [29]) or individual [30] opinion scores of beauty, when trained to do so. Features extracted by DCNNs has permitted the elaboration of feature-based metrics used to unravel the determinants of beauty [31,32]. One study has further shown that artistic and non-artistic images can be distinguished by variance of neuronal activations across layers of a DCNN [26]. Although the authors did not explicitly refer to DCNNs as models of human vision, the variance of neuronal activations describes a state of the visual system while processing an image and is thus a processing-based metric. Variance, however, is only loosely related to sparsity, and has no direct link to processing efficiency [33]. The use of DCNNs to measure processing efficiency and model fluency thus remains to be investigated.

In this work, analyzing images of various objects, we test the hypothesis that their beauty can be predicted from the sparsity of their DCNN activations. We compared results when analyzing various objects (non-artistic portrait photographs, figurative and abstract artistic paintings) and associated judgments related to beauty (beauty itself, attractiveness, negative-positive emotion) in order to qualitatively explore how generalizable the link between sparsity and beauty is. Beauty is ubiquitous, and can potentially describe any stimulus from natural landscapes to abstract representations, the physical aspect of a face or the artistic merit of its representation [34,35]. The fluency theory posits a unity of the experience of beauty across these domains, driven by a common biological mechanism: the efficiency of information processing. Moreover, fluency can mediate the evaluation of beauty in all three brain activities involved in aesthetic experiences: perception, cognition and emotions [36–38]. However, while aesthetic experiences can be both positive or negative (e.g., implying sadness or vertigo; [39]), like other authors we consider beauty as a subset of only positive, hedonically marked aesthetic experiences [5]. This distinction helps recognize that some artworks, for example, can have high aesthetic value while being not beautiful. We thus predict that high sparsity,

describing elevated fluency, is associated with high scores of beauty across visual domains and judgements of beauty.

## Results

We measured the sparsity of artificial neuron activations triggered by images for which an empirical score of beauty or judgment related to beauty has been obtained by psychological in-lab or online experiments. These images belong to four publicly available datasets and represent photographic portraits (CFD [40] and SCUT-FBP5500 [41] datasets) or artistic paintings (abstract painting: MART [42]; paintings from various styles and epochs: JEN [43] datasets; see *Materials and methods*). We analyzed these four datasets separately. Each image was first processed by the standard DCNN VGG-16 [44] pre-trained on the ImageNet dataset [45]. VGG-16 includes 13 convolutions and 2 fully connected layers. ImageNet is a large dataset of 14 million images depicting about 20,000 categories including people, plants, animals, and human-made objects. Training a DCNN on such a large and varied dataset allows modeling a visual cortex that is not specialized to one specific task, in accordance with neurophysiological data [23,46]. For each image, we extracted the activations for each of the 15 studied layers of the network and calculated the sparsity of the distribution of these activations. Thus, for each image, we have 15 measures of sparsity, one per layer.

### Link between sparsity and beauty

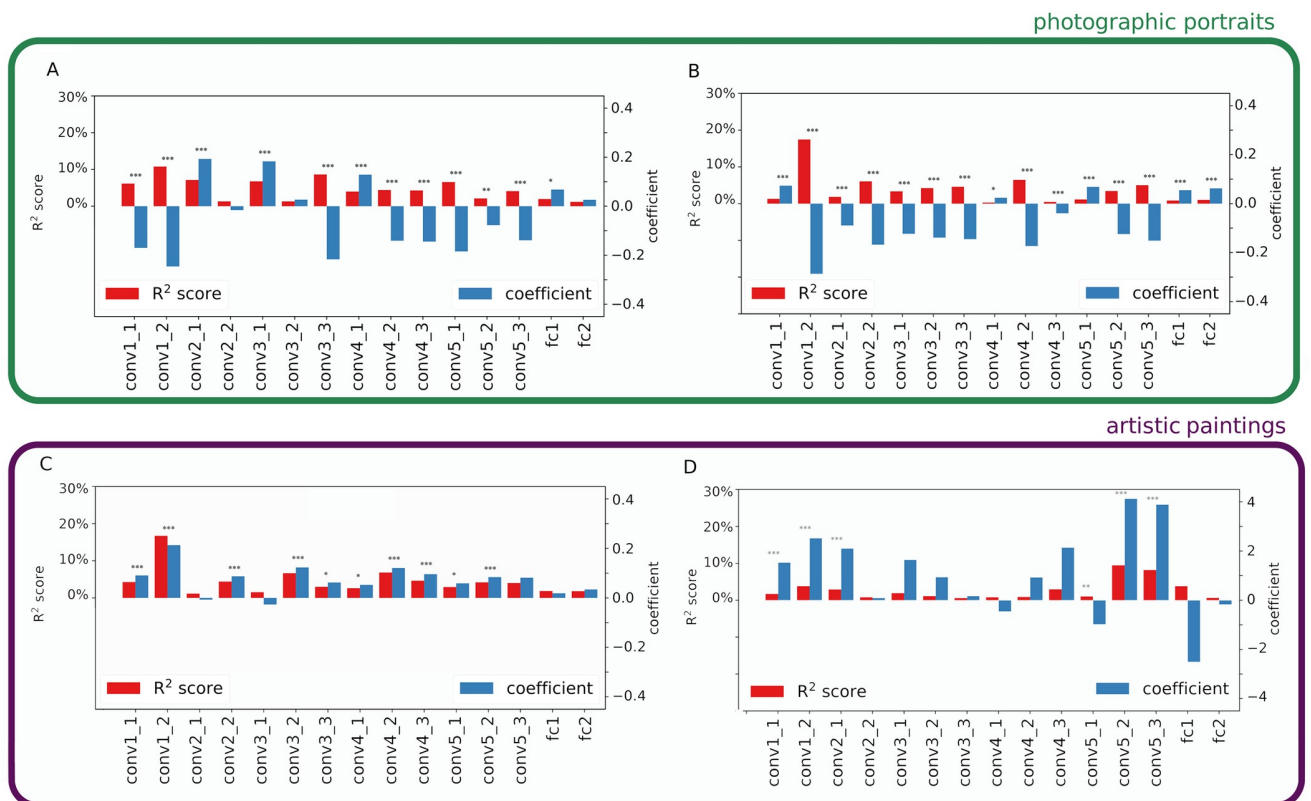
We studied whether beauty could be predicted from the sparsity of neuronal activations, and how this prediction varied with the visual domain (perception of photographic portraits vs. artistic paintings).

We investigated the different layers separately because each of them encodes different types of visual information and at different spatial scales. Early convolution layers encode information on simple and localized features (e.g., line segments, localized color contrasts), mid-level layers on moderately complex features like circles, and the last convolution and both fully connected layers more complex information (e.g., entire faces) spanning the whole image. For each layer separately, we estimated the variance in beauty scores (coefficient of determination  $R^2$  score) explained by sparsity using a linear regression model. We found a significant  $R^2$  score for all layers and all datasets ( $p < 10^{-6}$  for all models). However, the variance explained by sparsity varied strongly across both datasets (mean  $R^2$  across layers from 2,7% for JEN to 4,8% for CFD, Table 1), but also across layers for a given dataset (e.g., for SCUT-FBP5500 from  $R^2 = 0,3\%$  for layer conv4\_1 to  $R^2 = 17,5\%$  for layer conv1\_2; Fig 1). For each dataset, we observed at least one layer with a moderate  $R^2$  score (maximum: 11%, 18%, 8% and 9% for CFD, SCUT-FBP5500, JEN, MART, respectively). For both face datasets (CFD and SCUT-FBP5500)

**Table 1. Variance ( $R^2$  score) in beauty score explained by the sparsity and principal components of activations in VGG16.**  $R^2$  score are calculated between the ground truth and the predicted values of beauty scores by a multivariate Ridge linear regression model including sparsity of one layer (first column, mean of 15  $R^2$  scores, one per layer), sparsity of all layers (second column), principal component (PC) scores (explaining 80% of variance in activations) of one layer (third column: mean of 15  $R^2$  scores, one per layer), the scores of the first three PC of all layers (fourth column), or the scores of the first three PC and sparsity of all layers (fifth column) as predictors.  $R^2$  score was calculated using a 10-fold cross validation procedure. The four datasets (rows) were analyzed separately.

	mean $R^2$ (one model per layer, sparsity only)	$R^2$ (one model for all layers, sparsity only)	mean $R^2$ (one model per layer, PC scores only)	$R^2$ (one model for all layers, PC scores only)	$R^2$ (one model for all layers, PC scores + sparsity)
CFD	4.8%	24.7%	1.6%	4.2%	19.8%
SCUT-FBP5500	3.9%	28.3%	0.8%	4.2%	29.1%
MART	4.4%	25.7%	17.1%	46.4%	47.6%
JEN	2.7%	13.9%	0.7%	2.9%	14.2%

<https://doi.org/10.1371/journal.pcbi.1011703.t001>



**Fig 1. Explained variance in beauty score ( $R^2$  score) and coefficients of the sparsity of activations for each layer of VGG16.** A: CFD dataset. B: SCUT-FBP5500 dataset. C: MART dataset. D: JEN dataset. Sparsity is measured by the Gini index of the vectorized activations for each convolution and fully connected layers of VGG16 pre-trained on ImageNet. Coefficients were calculated by fitting univariate linear regression models fitted with a 10-fold cross-validation, and the  $R^2$  score was calculated between the ground truth and the predicted values. One, two or three asterisks indicate p-value smaller than 0.05, 0.01 and 0.001, respectively.

<https://doi.org/10.1371/journal.pcbi.1011703.g001>

and for the abstract paintings dataset (MART), the  $R^2$  score was the highest for the second convolution layer (conv1\_2). For the figurative paintings dataset (JEN), the  $R^2$  score was the highest for the two last convolution layers. This means that for faces and abstract paintings (CFD, SCUT-FBP5500, MART), people's perception of beauty is most influenced by the efficiency of processing texture or local color contrasts, while for figurative paintings, the efficiency of processing objects as a whole, and the spatial arrangement of their parts, is more important.

We also examined the coefficients estimated by this multivariate model (blue bars in Fig 1). Some of the layers with a significant effect of sparsity have positive coefficients while others have negative ones. Importantly, even within a broad portion of the network (early vs. mid vs. deep layers), the sign can vary from positive to negative, especially for face datasets (CFD, SCUT-FBP5500). This puzzling result indicates that from one layer to the next, high scores of beauty could be explained by either very efficient or inefficient neural codes.

We then investigated the link between beauty and the efficiency of information processing across the whole visual system. Indeed, previous studies in psychology have provided evidence that the ease of processing information at each stage of the visual system (e.g., as measured by detection time in the early stages, and as recognition performance in the later stages) triggers micro-experiences of fluency associated to each of these stages, and that these micro-experiences aggregate into one global sensation of fluency [5,47].

To model the aggregation of processing efficiency at every layer, we fitted a single multivariate (one measure of sparsity per layer) linear regression model to beauty scores. Because the model includes fifteen non-independent predictors (see S1 Fig), we constrained the optimization by applying a L2-norm penalty to coefficient estimates (i.e., Ridge regression model). We calculated the adjusted  $R^2$  score using a 10-fold cross validation procedure. We found a significant ( $p < 10^{-6}$ )  $R^2$  score for all four datasets. Moreover, sparsity explained approximately 25% of the variance in beauty scores for the two photographic portrait databases (CFD, SCUT-FBP5500) and for the abstract painting database (MART; Table 1, column " $R^2$  (one model for all layers, sparsity only)"). With figurative paintings, sparsity explained a substantial but lower fraction of variance in beauty scores ( $R^2 = 14\%$  for JEN; Table 1, column " $R^2$  (one model for all layers, sparsity only)"). Eventually, although a simple measure of sparsity averaged over DCNN layers does not explain much of the variation in beauty (Table 1, column " $mean R^2$  (one model per layer, sparsity only)"), a multivariate model including all layers predicts it with relatively high  $R^2$  scores. Yet, postulating that the multivariate regression models how fluency micro-experiences aggregate, our results show that this aggregation could be more complex than the mere sum of micro-experiences and may rather involve top-down controls weighting their importance in different stages of information processing.

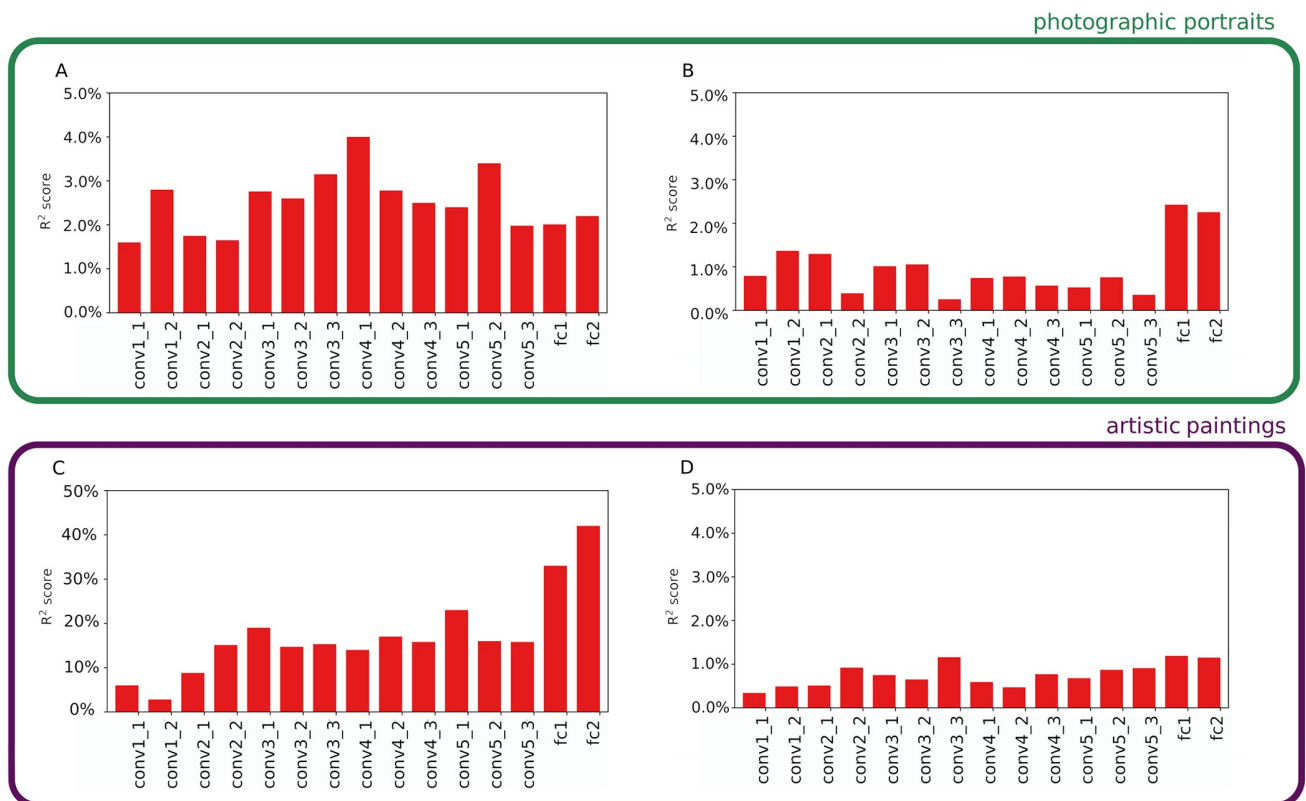
### Comparison with a feature-based model

In order to compare our processing-based approach to predict beauty to a traditional feature-based approach, we studied to which extent the neuronal activations themselves, rather than the sparsity of their distribution, can explain variation in beauty scores (following the method described by Iigaya [30]). In a DCNN, each artificial neuron describes one feature in a given region of the image, through its weighted connections to previous neurons. The magnitude of its activation indicates the importance of that specific feature in that region of the image. We then used neuronal activations as predictors in multivariate linear ridge regression models, fitting one model per layer at a time. However, because there were too many activations for fitting the models (the first convolution layer includes 3,211,264 activations), we reduced the number of activations using a principal component analysis (PCA), keeping only those components that explain a total of 80% of variance in activations (between 125 and 2,162 principal components depending on layers and datasets; see Supp Inf.). For all layers and all datasets, the  $R^2$  score was significant ( $p < 10^{-6}$  for all models), indicating that features, from simple and localized to complex and holistic, do explain variation in beauty scores for both face images and artistic paintings. However, activations explained a low fraction of variance in beauty scores, with  $R^2$  score never exceeding 4% (Fig 2). One major exception is for the abstract painting dataset MART, with  $R^2$  score ranging from 6% to 42%.

To analyze how encoded features could predict beauty at the entire network level, as for sparsity we used a single multivariate Ridge regression model using the PC scores of all layers. However, to keep the model tractable, we considered only the first three components of each layer. Again, except for MART, we found that the first three principal components explain a relatively low fraction of variance in beauty ( $R^2$  score  $< 5\%$ ; Table 1, column " $R^2$  (one model for all layers, PC scores only)"), in line with the previous result obtained when using all principal components that explain 80% of the variance in activations.

Last, we analyzed the benefits of combining both feature-based and processing-based metrics to predict beauty by fitting one multivariate ridge regression model per dataset, including the first three principal components of the PCA applied to activations plus sparsity for each layer, and considering all layers in the same model (i.e., 60 predictors). The  $R^2$  score was always significant ( $p < 10^{-6}$ ) and varied from 14% for the figurative paintings (JEN dataset) to up to





**Fig 2. Variance in beauty score ( $R^2$  score) explained by the principal components of activations in different layers of VGG16.** A: CFD dataset. B: SCUT-FBP5500 dataset. C: MART dataset. D: JEN dataset. The principal components are those explaining 80% of the total variance of the activations for each convolution and fully connected layers of VGG16 pre-trained on ImageNet. The  $R^2$  score was calculated between the ground truth and the predicted values by a multiple ridge linear regression model fitted with a 10-fold cross-validation procedure.

<https://doi.org/10.1371/journal.pcbi.1011703.g002>

48% for abstract paintings (Table 1, column “ $R^2$  (one model for all layers, PC scores + sparsity)”). Comparing these results to those with the model including sparsity only (Table 1, column “mean  $R^2$  (one model per layer, sparsity only)”), the increase in explained variance from adding the feature-based metrics is marginal (SCUT-FBP5500: +0.8%; JEN: +0.2%), or even negative (CFD: -4.9%). Only for the abstract paintings, considering features in addition to processing efficiency significantly increased the ability of the model to predict beauty scores (MART: +19.6%). Comparing these results to those with the model including PC scores only (Table 1, column “ $R^2$  (one model for all layers, PC scores only)”), considering sparsity increased the predictive capacities of the model by a factor 5 to 7. For the abstract paintings, considering sparsity only marginally increased the explained variance (MART: +1.2%).

## Discussion

It has been widely acknowledged that beauty results from the interaction between a stimulus and a beholder, however, the neuronal mechanisms describing this particular interaction have remained unknown. Here, we hypothesized that the way information is processed (processing-based aesthetics), more than the type of information being processed (feature-based aesthetics), drives the evaluation of beauty. We thus investigated whether visual beauty could be predicted from the sparsity of neuronal activations in an artificial deep convolutional neural network (DCNN), a statistic characterizing the efficiency of processing visual stimuli that is

independent of the features describing these stimuli. Our results clearly demonstrate the importance of considering the efficiency of information processing to predict beauty. Moreover, we found that for some categories of stimuli (faces and figurative artistic paintings), information processing itself likely contributes more than the processed features for explaining variation in perceived beauty. However, we also found that processing efficiency was either positively or negatively correlated with beauty, depending on which layer of the DCNN was considered.

Although demonstrated here with the help of artificial intelligence, the idea that the processing of stimuli *per se* is an important determinant of beauty is an old one in the field of aesthetics. The German philosopher Immanuel Kant, for example, proposed that beauty is not contained in the object, but is rather an effect of the state of mind of the subject, brought about by the object [48]. The theorist of literature Gérard Genette also famously claimed that “it is not the object that makes the relation aesthetic but the relation that makes the object aesthetic” [49]. Our results, showing that sparsity can explain variation in beauty across different visual domains and components of beauty (i.e. when evaluating the physical beauty of faces, or the artistic beauty of abstract or figurative paintings), offer strong support to the interactionist paradigm of beauty and suggest a common biological mechanism underlying all evaluations of beauty.

One previous study proposed to quantify fluency through statistical typicality [50], based on observations that more typical stimuli are also more easily processed [51]. Statistical typicality is related to the psychological concept of familiarity, the manipulation of which has been pivotal in supporting the fluency theory of aesthetics (e.g., [36]). Typicality was measured as the likelihood of a stimulus in the latent space of an encoder, given a reference distribution. Using CFD as the reference distribution, the authors found that facial typicality explained 15% of variance in facial attractiveness. Crucially, sparsity also captures the concept of statistical typicality because, during training, the weights of a DCNN are tuned to the statistical distribution of features of the training dataset (in our case, ImageNet), such that images with a similar distribution are sparsely encoded during inference [52]. However, while typicality specifically refers to familiarity, sparsity more explicitly characterizes the use of neuronal resources, that is, the efficiency of information processing *per se*, a presumably important component of fluency [8]. Regardless of the level of overlap between typicality and sparsity, both [34] and this study considered a static perceptual system. In contrast, a classical approach to study fluency in psychology is to investigate how the system varies when manipulating fluency, for example using matching prime (see [36]). In the future, it would thus be interesting to apply similar dynamic approaches to DCNNs, for example by analyzing how progressively tuning the DCNN to features that are specific to the target stimuli increases the ability of sparsity to explain beauty.

Comparing the predictive capacity of a model including sparsity only with a model including both sparsity and features suggests that information processing could be more important than features in determining beauty. One exception is for the abstract paintings MART dataset, for which features explained more variation in beauty scores than sparsity did, and sparsity did not improve prediction. The fact that sparsity did not significantly improve the model with PC scores only, while it could alone explain 25.7% of variance of beauty scores, suggests that, for abstract paintings, most of the information conveyed by processing efficiency is redundant with information conveyed by features. Given that the negative-positive emotions were evaluated in the abstract paintings dataset, one explanation to the difference between this dataset and other datasets could be that the hedonic marking of efficient information processing in general is not mediating the link between beauty and emotions. This, however, contradicts previous accounts suggesting at least a mild effect of fluency on aesthetic emotions (for a

review, see [53]). For this dataset, beauty was scored by lay people who thus based their judgment on the design of stimuli only, and not on representations encoded in as in figurative paintings. Our result could thus rather suggest that information processing influences aesthetics only when stimuli elicit interpretation or meaning assignment (see also [54]).

The relative importance of sparsity and features in predicting beauty should be interpreted with caution, however. Indeed, the principal component analysis is a commonly used method to describe features (*e.g.*, [30,55]), but it is reductive by construction. Unfortunately, accounting for the full variation in features inevitably poses limitations, either technical or biological. The use of a PCA to reduce the dimensionality of the feature space is an example of technical limitation imposed by statistical regression modeling. To circumvent this limitation, other studies have trained DCNN to directly predict beauty scores, obtaining remarkably high  $R^2$  scores (up to 90% in [56]). While these studies highlight that features can theoretically encode most of the information about beauty, in practice, they are not biologically realistic because, contrary to DCNN, the visual cortex as a whole has been shaped by evolution and development to perform many tasks, not to predict beauty only. More work is clearly needed both to quantify the relative importance of processing fluency and stimulus features in predicting beauty, and to understand the factors influencing this relative importance.

Interestingly, for a given dataset, the most predictive layers differ when considering either information processing or features. For MART and SCUT-FBP5500, for example, beauty is best predicted by the last two layers, Fc1 and Fc2 (describe whole-image features) when considering PCA features, and by the first convolutional layers (describe textures and local color contrasts) when considering sparsity. The dissociation between processing-based and feature-based contributions to beauty highlights the richness of the aesthetic experience and myriad possibilities that naturally or culturally shaped communication signals have to increase their attractiveness.

Although our results validate our main hypothesis that beauty can be explained in part from information processing, they reject the prediction that high sparsity is associated with high scores of beauty. Indeed, we found that depending on layers and datasets, high beauty scores could be associated with either highly efficient, or the opposite, highly energy-demanding coding (see Fig 1). This result contradicts the existence of a universal positive correlation between processing efficiency and beauty, and thus the original (and most frequently cited) formulation of the fluency theory of aesthetics. Yet, a recent study in psychology has also found that beauty could be negatively correlated with fluency [57]. One explanation is that fluency would work as an amplifier, being positively correlated with beauty for stimuli with a positive emotional valence, and negatively correlated for negatively-valenced stimuli [58]. This is in line with the general idea that a core goal of perceptual processing is to disambiguate stimuli and thus enable the most appropriate behavior [59]. Another explanation is that relative, rather than absolute fluency, determines beauty [9], and thus that alternating between low and high fluency contributes to increasing beauty [60]. With suspense, or optical illusions, for example, pleasure arises from a sudden increase in information processing that was purposely blocked to amplify the feeling of fluency. In music, alternating between consonant and dissonant chord structures contributes to the pleasure of listening and exemplifies the more general tension-resolution hypothesis of aesthetic experience [61]. Whether similar phenomena occurring between processing stages within the information pathway is necessary to trigger an aesthetic experience remains to be investigated, but our data from artificial neural networks suggest that these hypotheses deserve investigation. In any case, when considering the different stages of visual perception, fluency appears to be linked to beauty in more complex ways than previously thought.

Finally, our study relies on both assumptions that DCNNs can model information processing, and that they are valuable tools to study complex mental phenomena. The first assumption has remained controversial [62] as several studies have highlighted differences, some of them significant, between DCNNs and biological vision in the underlying computation (e.g. [63,64]). For example, DCNNs rely on texture information more than shape in object recognition [65]. Yet for many neuroscientists, the emergence in DCNNs of the most fundamental properties observed in animal vision indicates that they are still powerful tools for modelling and understanding this perceptual modality [66,67]. Regarding the second assumption, we acknowledge that purely feed-forward models like VGG16 fall short in describing the full complexity of the judgment of beauty. However, despite top-down cognitive controls that likely influence this judgment, several studies have stressed that it can be made independently of any appeal to cognition [68–70]. Like previous studies showing that VGG16 models reasonably well how the human brain generally processes visual information (e.g., [19]), we concur that the simplicity and tractability of this DCNN can reveal overlooked important brain processes. In particular, it can open a new era in the centuries-old quest to explain and predict beauty.

## Materials and methods

### Datasets

We used four different, publicly available datasets consisting of images and associated mean opinion scores (MOS) of beauty (Table 2). The Chicago Face Database (CFD v3.0) includes 827 standardized photographic portraits (centered faces, identical outfits, identical camera settings) of American and Indian adults of both genders, from various self-reported ethnic origins [40]. Image size is 2,444x1,718 pixels. For each image, a MOS of attractiveness was obtained by averaging ratings from 1,087 American participants, along a Likert scale from 1 to 7, with 1 being not attractive at all and 7 extremely attractive. (Question R013 from CFD Codebook). See S2 Table for more information on the evaluators.

The SCUT-FBP5500 database (South China University of Technology—Facial Beauty Prediction) includes 5,500 photographic portraits of Asian and Caucasian adults of both genders [41]. Portrait images were retrieved from various sources and thus are not standardized. Image size is 350x350 pixels. Beauty MOS were obtained by averaging ratings from 60 Chinese volunteers, along a Likert scale from 1 to 5 with 1 being the least attractive and 5 the most attractive. See S2 Table for more information on the evaluators.

The MART database (Museum of Modern and Contemporary Art of Trento and Rovereto) includes images of 500 abstract paintings by different artists from Trento museum in Italy [42]. The width and height of images varies from 59 to 812, and 45 to 1,036 pixels, respectively. One hundred Italian laypersons were asked to rate their emotional response to each image along a Likert scale from 1 to 7, with 1 being the most negative and 7 the most positive. Previous studies have shown that beauty is mostly associated with positive emotions (e.g., [71]). See S2 Table for more information on the evaluators.

**Table 2. Description of image databases associated with a beauty score.**

Name	# Images	Type	Evaluated component of beauty
CFD	827	Portraits	Attractiveness
SCUT-FBP5500	5,500	Portraits	Beauty
MART	500	Abstract paintings	Negative-positive emotion
JEN	1,563	Representational paintings	Beauty

<https://doi.org/10.1371/journal.pcbi.1011703.t002>

Last, the JEN database (from the German town of Jena) contains 1,563 figurative artistic paintings by different artists representing different movements [43]. Image size is 1,456–30,000x1,351–23,803 pixels. For each image, beauty was evaluated by 134 observers along a Likert scale from 0 to 100, with 0 being not beautiful and 100 beautiful. See [S2 Table](#) for more information on the evaluators.

For all analyses, images were resampled to 224x224 pixels (input size of VGG16), and scores of beauty or their proxy were all standardized to the same scale varying between 0 (lowest attractiveness/beauty score) and 5 (highest attractiveness/beauty score).

### Encoding images with a convolutional neural network

All images were encoded using the VGG16 [44] deep convolutional neural network pretrained on the ImageNet dataset [45]. VGG16 includes 13 convolution layers and two fully connected layers. For each image, we thus extracted 15 encodings corresponding to the neuronal activations (after ReLU transformation) of these 15 different layers. For convolution layers, encodings are three-dimensional matrices of size  $H \times W \times C$ , with  $H$ ,  $W$  corresponding to the height and the width of the feature maps, respectively, and  $C$  to the number of feature maps (i.e., channels).  $H$  and  $W$  vary from 224 to 14, and  $C$  from 64 to 512, between the first and the last convolution layer. For both fully connected layers, encodings are one-dimensional vectors of size 4,096.

### Measuring sparsity

The efficiency of image processing at a given network layer was estimated as the activation sparsity, which measures inequity in the distribution of activations in this layer. One study evaluated the ability of various metrics of sparsity to satisfy the attributes desired to properly measure inequity of distribution [72]. The authors found that only the Gini index meets all the desired attributes. For example, kurtosis and  $L_1$ -norm, two other commonly used metrics of sparsity, fail to predict an increase in inequity when adding null values and extremely high values, respectively. The authors did not include the Treves-Rolls metric [73] in their comparison, despite its popular use in neuroscience. We thus used the Gini index and the Treves-Rolls metrics to measure sparsity. The two metrics yielded qualitatively similar results, thus only results with the Gini index are presented here.

For convolution layers, the  $H \times W \times C$  matrices of activations were first flattened into one-dimensional vectors. These vectors, and the vectors of fully connected layers were sorted in ascending order. The Gini index was calculated as:

$$G = \frac{\sum_{i=1}^n (2i - n - 1)x_i}{n \sum_{i=1}^n x_i}$$

with  $n$  the number of activations in the layer (i.e., vector length),  $i$  the index and  $x_i$  the activation at the index. Image encodings and measurements of sparsity were performed using Python 3.9.5.

### Statistical analyses

The four datasets were studied separately. For each layer, measurements of sparsity were z-transformed prior to statistical analyses. We first conducted simple linear regression models to analyze the correlation between the empirical scores of beauty (response variable) and the sparsity measured for each layer (one model per layer). Then, we analyzed the ability to predict beauty from sparsity measurements of all layers with a single multivariate linear regression

model and a Ridge penalization, that included 15 explanatory variables (a separate sparsity measurement for each of the 15 layers). The strength of the Ridge penalization (Lambda parameter) was set by cross-validation.

Last, we used linear Ridge regression models to analyze the ability of image features to predict beauty scores. However, encodings have too many dimensions to be included in a regression model. For example, in the first convolution layer, the encoding corresponds to  $H^*W^*C = 3,211,264$  features. We thus first reduced the dimensionality of encodings by applying a PCA (one PCA per layer per dataset), keeping the principal components explaining 80% of variance (see S1 Table). PC (principal components) scores were then z-transformed and included in the ridge regression model as explanatory variables. We performed one Ridge regression for each layer separately. In addition, we performed one global Ridge regression including the first three principal components of each layer (model with 45 explanatory variables), and another global Ridge regression including the first three principal components and the sparsity measurement of each layer (model with 60 explanatory variables). For all linear regression and linear Ridge models, we performed a 10-fold cross-validation, repeated 100 times to ensure that the results were stable, calculating the coefficient of determination ( $R^2$  score for univariate models and adjusted  $R^2$  score for multivariate models) between predicted and empirical beauty scores for each test fold, and then averaging the  $R^2$  score over the 10 folds and then over the 100 repetitions. Statistical analyses were performed using R software v4.2.1 [74]. We used the *glmnet* method of package *glmnet* [75] for the Ridge regression models, and the *caret* package for the cross validation [76].

## Supporting information

**S1 Fig. Pearson correlation matrix of the sparsity of activations in different layers of VGG16.** A: CFD dataset. B: SCUT-FBP5500 dataset. C: MART dataset. D: JEN dataset. (TIF)

**S2 Fig. Beauty score explained by the Gini index for the layer with the highest  $R^2$  score per database.** A: CFD dataset (second convolution layer of the first block). B: SCUT-FBP5500 dataset (second convolution layer of the first block). C: MART dataset (second convolution layer of the first block). D: JEN dataset (second convolution layer of the fifth block). The blue trend line corresponds to the predictions provided by the model. (TIF)

**S1 Table. Number of PCA components that explained 80% of the total variance of activations for each database and each layer of VGG16.** A: CFD dataset. B: SCUT-FBP5500 dataset. C: MART dataset. D: JEN dataset. (DOCX)

**S2 Table. Description of evaluators from image datasets.** (DOCX)

## Acknowledgments

We are grateful to Melvin Bardin for his help with high-performance computing.

## Author Contributions

**Conceptualization:** Nicolas M. Dibot, Tamra C. Mendelson, Julien P. Renoult.

**Formal analysis:** Nicolas M. Dibot.

**Investigation:** Julien P. Renoult.

**Methodology:** Nicolas M. Dibot, Sonia Tieo, Julien P. Renoult.

**Project administration:** Julien P. Renoult.

**Supervision:** William Puech, Julien P. Renoult.

**Writing – original draft:** Nicolas M. Dibot, Julien P. Renoult.

**Writing – review & editing:** Nicolas M. Dibot, Sonia Tieo, Tamra C. Mendelson, William Puech, Julien P. Renoult.

## References

1. Brachmann A, Redies C. Computational and Experimental Approaches to Visual Aesthetics. *Front Comput Neurosci*. 2017; 11: 102. <https://doi.org/10.3389/fncom.2017.00102> PMID: 29184491
2. Balietti S. The human quest for discovering mathematical beauty in the arts. *Proc Natl Acad Sci*. 2020; 117: 27073–27075. <https://doi.org/10.1073/pnas.2018652117> PMID: 33097664
3. Jacobsen T, Höfel L. Aesthetic judgments of novel graphic patterns: analyses of individual judgments. *Percept Mot Skills*. 2002; 95: 755–766. <https://doi.org/10.2466/pms.2002.95.3.755> PMID: 12509172
4. McManus IC. The aesthetics of simple figures. *Br J Psychol*. 1980; 71: 505–524. <https://doi.org/10.1111/j.2044-8295.1980.tb01763.x> PMID: 7437674
5. Reber R, Schwarz N, Winkielman P. Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personal Soc Psychol Rev*. 2004; 8: 364–382. [https://doi.org/10.1207/s15327957pspr0804\\_3](https://doi.org/10.1207/s15327957pspr0804_3) PMID: 15582859
6. Ishizu T, Zeki S. Toward A Brain-Based Theory of Beauty. *PLOS ONE*. 2011; 6: 1–10. <https://doi.org/10.1371/journal.pone.0021852> PMID: 21755004
7. Winkielman P, Schwarz N, Fazendeiro T, Reber R. The Hedonic marking of Processing Fluency: implications for evaluative judgment. *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. Psychology Press; 2003.
8. Renoult JP, Mendelson TC. Processing bias: extending sensory drive to include efficacy and efficiency in information processing. *Proc R Soc B Biol Sci*. 2019; 286: 20190165. <https://doi.org/10.1098/rspb.2019.0165> PMID: 30940061
9. Muth C, Carbon C-C. The aesthetic aha: on the pleasure of having insights into Gestalt. *Acta Psychol (Amst)*. 2013; 144: 25–30. <https://doi.org/10.1016/j.actpsy.2013.05.001> PMID: 23743342
10. Oppenheimer DM. The secret life of fluency. *Trends Cogn Sci*. 2008; 12: 237–241. <https://doi.org/10.1016/j.tics.2008.02.014> PMID: 18468944
11. Nisbett RE, Wilson TD. Telling more than we can know: Verbal reports on mental processes. *Psychol Rev*. 1977; 84: 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>
12. Mayer S, Landwehr J. Quantifying visual aesthetics based on processing fluency theory: Four algorithmic measures for antecedents of aesthetic preferences. *Psychol Aesthet Creat Arts*. 2018; 12: 399–431. <https://doi.org/http%3A//dx.doi.org/10.1037/aca0000187>
13. Redies C. A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spat Vis*. 2008; 21: 97–117. <https://doi.org/10.1163/156856808782713780>
14. Olshausen BA, Field DJ. Sparse coding of sensory inputs. *Curr Opin Neurobiol*. 2004; 14: 481–487. <https://doi.org/10.1016/j.conb.2004.07.007> PMID: 15321069
15. Winkielman P, Huber DE, Kavanagh L, Schwarz N. Fluency of consistency: When thoughts fit nicely and flow smoothly. *Cognitive Consistency: A Fundamental Principle in Social Cognition*. 2012. pp. 89–111.
16. Renoult JP, Bovet J, Raymond M. Beauty is in the efficient coding of the beholder. *R Soc Open Sci*. 2016; 3: 160027. <https://doi.org/10.1098/rsos.160027> PMID: 27069668
17. Holzeleitner IJ, Lee AJ, Hahn AC, Kandrik M, Bovet J, Renoult JP, et al. Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *J Exp Psychol Hum Percept Perform*. 2019; 45: 1589–1595. <https://doi.org/10.1037/xhp0000685> PMID: 31556686
18. Hibbard PB, O'Hare L. Uncomfortable images produce non-sparse responses in a model of primary visual cortex. *R Soc Open Sci*. 2: 140535. <https://doi.org/10.1098/rsos.140535> PMID: 26064607
19. Lindsay GW. Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *J Cogn Neurosci*. 2020; 1–15. [https://doi.org/10.1162/jocn\\_a\\_01544](https://doi.org/10.1162/jocn_a_01544) PMID: 32027584

20. Kriegeskorte N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annu Rev Vis Sci.* 2015; 1: 417–446. <https://doi.org/10.1146/annurev-vision-082114-035447> PMID: 28532370
21. Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, Bethge M, et al. Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLOS Comput Biol.* 2019; 15: e1006897. <https://doi.org/10.1371/journal.pcbi.1006897> PMID: 31013278
22. Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci Rep.* 2016; 6: 27755. <https://doi.org/10.1038/srep27755> PMID: 27282108
23. Güçlü U, Gerven MAJ van. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J Neurosci.* 2015; 35: 10005–10014. <https://doi.org/10.1523/JNEUROSCI.5023-14.2015> PMID: 26157000
24. Khaligh-Razavi S-M, Kriegeskorte N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Comput Biol.* 2014; 10: e1003915. <https://doi.org/10.1371/journal.pcbi.1003915> PMID: 25375136
25. Saraee E, Jalal M, Betke M. Visual complexity analysis using deep intermediate-layer features. *Comput Vis Image Underst.* 2020; 195: 102949. <https://doi.org/10.1016/j.cviu.2020.102949>
26. Brachmann A, Barth E, Redies C. Using CNN features to better understand what makes visual artworks special. *Front Psychol.* 2017; 8: 830. <https://doi.org/10.3389/fpsyg.2017.00830> PMID: 28588537
27. Cetinic E, She J. Understanding and Creating Art with AI: Review and Outlook. *ACM Trans Multimed Comput Commun Appl.* 2022; 18: 66:1–66:22. <https://doi.org/10.1145/3475799>
28. Sandoval C, Pirogova E, Lech M. Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings. *IEEE Access.* 2019; 7: 41770–41781.
29. Lin L, Liang L, Jin L. Regression Guided by Relative Ranking Using Convolutional Neural Network (R<sup>3</sup>CNN) for Facial Beauty Prediction. *IEEE Trans Affect Comput.* 2022; 13: 122–134. <https://doi.org/10.1109/TAFFC.2019.2933523>
30. Iigaya K, Yi S, Wahle IA, Tanwisuth K, O'Doherty JP. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nat Hum Behav.* 2021; 5: 743–755. <https://doi.org/10.1038/s41562-021-01124-6> PMID: 34017097
31. Hosu V, Goldlucke B, Saupé D. Effective Aesthetics Prediction with Multi-level Spatially Pooled Features. *arXiv*; 2019. <https://doi.org/10.48550/arXiv.1904.01382>
32. Gan J, Jiang K, Tan H, He G. Facial Beauty Prediction Based on Lighted Deep Convolution Neural Network with Feature Extraction Strengthened. *Chin J Electron.* 2020; 29: 312–321. <https://doi.org/10.1049/cje.2020.01.009>
33. Tolhurst DJ, Smyth D, Thompson ID. The Sparseness of Neuronal Responses in Ferret Primary Visual Cortex. *J Neurosci.* 2009; 29: 2355–2370. <https://doi.org/10.1523/JNEUROSCI.3869-08.2009> PMID: 19244512
34. Saito Y. *Everyday Aesthetics.* Oxford University Press; 2007. <https://doi.org/10.1093/acprof:oso/9780199278350.001.0001>
35. Schulz K, Hayn-Leichsenring GU. Face Attractiveness versus Artistic Beauty in Art Portraits: A Behavioral Study. *Front Psychol.* 2017;8. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02254>
36. Reber R, Winkielman P, Schwarz N. Effects of perceptual fluency on affective judgments. *Psychol Sci.* 1998; 9: 45–48.
37. Belke B, Leder H, Strobach T, Carbon C-C. Cognitive fluency: High-level processing dynamics in art appreciation. *Psychol Aesthet Creat Arts.* 2010; 4: 214–222. <https://doi.org/10.1037/a0019648>
38. Kuchinke L, Trapp S, Jacobs AM, Leder H. Pupillary responses in art appreciation: Effects of aesthetic emotions. *Psychol Aesthet Creat Arts.* 2009; 3: 156–163. <https://doi.org/10.1037/a0014464>
39. Brady E. *The Sublime in Modern Philosophy: Aesthetics, Ethics, and Nature.* Cambridge University Press; 2013. <https://doi.org/10.1017/CBO9781139018098>
40. Ma DS, Correll J, Wittenbrink B. The Chicago face database: A free stimulus set of faces and norming data. *Behav Res Methods.* 2015; 47: 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5> PMID: 25582810
41. Xie D, Liang L, Jin L, Xu J, Li M. SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception. 2015. <https://doi.org/10.48550/arXiv.1511.02459>
42. Yanulevskaya V, Uijlings J, Bruni E, Sartori A, Zamboni E, Bacci F, et al. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. *Proceedings of the 20th*



- ACM international conference on Multimedia. New York, NY, USA: Association for Computing Machinery; 2012. pp. 349–358. <https://doi.org/10.1145/2393347.2393399>
43. Amirshahi SA, Hayn-Leichsenring GU, Denzler J, Redies C. JenAesthetics Subjective Dataset: Analyzing Paintings by Subjective Scores. In: Agapito L, Bronstein MM, Rother C, editors. Computer Vision—ECCV 2014 Workshops. Cham: Springer International Publishing; 2015. pp. 3–19. [https://doi.org/10.1007/978-3-319-16178-5\\_1](https://doi.org/10.1007/978-3-319-16178-5_1)
  44. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015. <https://doi.org/10.48550/arXiv.1409.1556>
  45. Fei-Fei L, Deng J, Li K. ImageNet: Constructing a large-scale image database. *J Vis.* 2009; 9: 1037. <https://doi.org/10.1167/9.8.1037>
  46. Peterson JC, Abbott JT, Griffiths TL. Evaluating (and Improving) the Correspondence Between Deep Neural Networks and Human Representations. *Cogn Sci.* 2018; 42: 2648–2669. <https://doi.org/10.1111/cogs.12670> PMID: 30178468
  47. Wurtz P, Reber R, Zimmermann TD. The feeling of fluent perception: a single experience from multiple asynchronous sources. *Conscious Cogn.* 2008; 17: 171–184. <https://doi.org/10.1016/j.concog.2007.07.001> PMID: 17697788
  48. Kant I. Immanuel Kant: Kritik der Urteilskraft. Akademie Verlag; 1790.
  49. Canvat K. L'œuvre de l'art. T.2. La relation esthétique. Gérard Genette. Seuil, coll. «Poétique », Paris, 1997. *Lett AIRDF.* 1999; 24: 26–26.
  50. Ryali CK, Goffin S, Winkelman P, Yu AJ. From likely to likable: The role of statistical typicality in human social assessment of faces. *Proc Natl Acad Sci.* 2020; 117: 29371–29380. <https://doi.org/10.1073/pnas.1912343117> PMID: 33229540
  51. Winkelman P, Halberstadt J, Fazendeiro T, Catty S. Prototypes are attractive because they are easy on the mind. *Psychol Sci.* 2006; 17: 799–806. <https://doi.org/10.1111/j.1467-9280.2006.01785.x> PMID: 16984298
  52. Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. pp. 2892–2900. <https://doi.org/10.1109/CVPR.2015.7298907>
  53. Armstrong T, Detweiler-Bedell B. Beauty as an emotion: The exhilarating prospect of mastering a challenging world. *Rev Gen Psychol.* 2008; 12: 305–329. <https://doi.org/10.1037/a0012558>
  54. Graf LKM, Landwehr JR. A dual-process perspective on fluency-based aesthetics: the pleasure-interest model of aesthetic liking. *Personal Soc Psychol Rev Off J Soc Personal Soc Psychol Inc.* 2015; 19: 395–410. <https://doi.org/10.1177/1088868315574978> PMID: 25742990
  55. Garg I, Panda P, Roy K. A Low Effort Approach to Structured CNN Design Using PCA. *IEEE Access.* 2020; 8: 1347–1360. <https://doi.org/10.1109/ACCESS.2019.2961960>
  56. Bougourzi F, Dornaika F, Taleb-Ahmed A. Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. *Knowl-Based Syst.* 2022; 242: 108246. <https://doi.org/10.1016/j.knosys.2022.108246>
  57. Landwehr JR, Eckmann L. The nature of processing fluency: Amplification versus hedonic marking. *J Exp Soc Psychol.* 2020; 90: 103997. <https://doi.org/10.1016/j.jesp.2020.103997>
  58. Albrecht S, Carbon C-C. The Fluency Amplification Model: fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychol (Amst).* 2014; 148: 195–203. <https://doi.org/10.1016/j.actpsy.2014.02.002> PMID: 24603044
  59. Carbon C-C, Albrecht S. The Fluency Amplification Model supports the GANE principle of arousal enhancement. *Behav Brain Sci.* 2016; 39: e204. <https://doi.org/10.1017/S0140525X15001752> PMID: 28347372
  60. Schaeffer J-M. L'expérience esthétique. Editions Gallimard; 2015.
  61. Lehne M, Koelsch S. Tension-resolution patterns as a key element of aesthetic experience: psychological principles and underlying brain mechanisms. *Art Aesthet Brain.* 2014;545.
  62. Geirhos R, Meding K, Wichmann FA. Beyond accuracy: quantifying trial-by-trial behaviour of CNNs and humans by measuring error consistency. *Advances in Neural Information Processing Systems.* Curran Associates, Inc.; 2020. pp. 13890–13902. Available: <https://proceedings.neurips.cc/paper/2020/hash/9f6992966d4c363ea0162a056cb45fe5-Abstract.html>
  63. Nonaka S, Majima K, Aoki SC, Kamitani Y. Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *iScience.* 2021; 24: 103013. <https://doi.org/10.1016/j.isci.2021.103013> PMID: 34522856

64. Lonnqvist B, Bornet A, Doerig A, Herzog MH. A comparative biology approach to DNN modeling of vision: A focus on differences, not similarities. *J Vis.* 2021; 21: 17. <https://doi.org/10.1167/jov.21.10.17> PMID: 34551062
65. Baker N, Lu H, Erlikhman G, Kellman PJ. Local features and global shape information in object classification by deep convolutional neural networks. *Vision Res.* 2020; 172: 46–61. <https://doi.org/10.1016/j.visres.2020.04.003> PMID: 32413803
66. Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, et al. A deep learning framework for neuroscience. *Nat Neurosci.* 2019; 22: 1761–1770. <https://doi.org/10.1038/s41593-019-0520-2> PMID: 31659335
67. Saxe A, Nelli S, Summerfield C. If deep learning is the answer, what is the question? *Nat Rev Neurosci.* 2021; 22: 55–67. <https://doi.org/10.1038/s41583-020-00395-8> PMID: 33199854
68. Kunst-Wilson WR, Zajonc RB. Affective Discrimination of Stimuli That Cannot Be Recognized. *Science.* 1980; 207: 557–558. <https://doi.org/10.1126/science.7352271> PMID: 7352271
69. Janiszewski C. Preattentive Mere Exposure Effects. *J Consum Res.* 1993; 20: 376–392. <https://doi.org/10.1086/209356>
70. Murphy ST, Zajonc RB. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *J Pers Soc Psychol.* 1993; 64: 723–739. <https://doi.org/10.1037//0022-3514.64.5.723> PMID: 8505704
71. Schindler I, Hosoya G, Menninghaus W, Beermann U, Wagner V, Eid M, et al. Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE.* 2017; 12: e0178899. <https://doi.org/10.1371/journal.pone.0178899> PMID: 28582467
72. Hurley N, Rickard S. Comparing Measures of Sparsity. *IEEE Trans Inf Theory.* 2009; 55: 4723–4741. <https://doi.org/10.1109/TIT.2009.2027527>
73. Rolls ET, Treves A. The neuronal encoding of information in the brain. *Prog Neurobiol.* 2011; 95: 448–490. <https://doi.org/10.1016/j.pneurobio.2011.08.002> PMID: 21907758
74. Ihaka R. and Gentleman R. (1993) The R Project for Statistical Computing.—References— Scientific Research Publishing. [cited 28 Feb 2023]. Available: [https://www.scirp.org/\(S\(vtj3fa45qm1ean45%20vffcz55\)\)/reference/referencespapers.aspx?referenceid=2697689](https://www.scirp.org/(S(vtj3fa45qm1ean45%20vffcz55))/reference/referencespapers.aspx?referenceid=2697689)
75. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, et al. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. 2022. Available: <https://CRAN.R-project.org/package=glmnet>
76. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;28. <https://doi.org/10.18637/jss.v028.i05>



# **DISCUSSION GÉNÉRALE**



# Préambule à la discussion générale

# 7

Les travaux réalisés dans cette thèse ont permis d'apporter des éléments de réponses à la fois à des questions scientifiques, méthodologiques et techniques. Au-delà de tester des hypothèses, nous avons ouvert des pistes de réflexion sur des perspectives plus larges. Nous discuterons en partie 1 d'en quoi les résultats obtenus pris dans leur ensemble permettent d'éclairer l'origine des préférences. En partie 2, nous nous focaliserons sur les perspectives et les limites de notre approche expérimentale, ainsi que sur les limites des expérimentations basées sur le temps d'observation en général. Enfin, en partie 3, nous proposerons une discussion sur l'intérêt de certains aspects de l'IA en écologie et en évolution.



Nous avons vu que les préférences, qu'elles portent sur des visages ou non, peuvent s'expliquer par deux théories distinctes.

Premièrement, un signal visuel peut être un indicateur de qualité. Ainsi, l'apparition d'une préférence pour ce signal se justifie par l'intérêt d'interagir avec l'individu qui possède le signal. Une telle interaction (coopération, reproduction etc) aurait alors un intérêt évolutif pour l'individu attiré par le signal. Par exemple, chez les humains, nous avons vu en introduction que des visages de femmes plus féminins peuvent être préférés car ils sont un indicateur de compétences maternelles et de fertilité. Deuxièmement, une préférence peut être latente et s'expliquer par un biais perceptuel, qui correspond à l'adaptation de système perceptif (visuel en ce qui concerne cette thèse) à des fonctions qui ne sont pas directement liées aux traits visés par la préférence. Par exemple, les prototypes, représentant la moyenne d'une catégorie, sont préférés à des stimuli atypiques, car il serait plus facile de s'en faire une image mentale. Des visages de femmes plus féminins peuvent être préférés car représentatifs de la moyenne des visages féminins.

Concernant l'attractivité pour la féminité, il est donc difficile de trancher pour l'une ou pour l'autre de ces théories, toutes deux validées par des résultats empiriques. S'éloigner du modèle humain permet ainsi de voir le problème sous un autre angle. Ainsi, chez les mandrills, nous avons montré expérimentalement que les mandrills mâles préfèrent les mandrills femelles moins féminins. Ce résultat renforce des observations réalisées sur des populations sauvages et indique que la préférence pour la féminité s'explique surtout par ce que ce trait est un indicateur de qualité, puisque la valence de cette préférence n'est pas commune à deux espèces de primates différentes, les humains et les mandrills. En effet, les femelles moins féminines sont plus éloignées de femelles féminines et donc typiques, invalidant ainsi l'hypothèse que l'origine de cette préférence soit un biais perceptuel basé sur la typicalité.

Les comportements des humains peuvent néanmoins offrir des pistes de réflexion sur les raisons de cette préférence. Ainsi, dans certains contextes, les femmes moins féminines peuvent être préférées car considérées comme plus dignes de confiance. Par exemple, une féminité élevée est perçue comme un risque accru d'infidélité [137] ou est associée à des préjugés comme une compétence moindre [138, 139].

Nous avons aussi traité la question de l'attractivité des visages sous un autre prisme que celui de la féminité : le prisme de la beauté. L'approche, très différente car basée sur la modélisation du système perceptif des mammifères avec des réseaux de neurones convolutifs, nous a permis de montrer que sous ce prisme, un biais perceptuel, la fluence, pouvait avoir un impact important mais néanmoins complexe sur la beauté. La théorie de la fluence postule que la beauté est provoquée par cette fluence, une sensation agréable déterminée par l'efficacité du traitement de l'information dans le système visuel. Pour autant, nous avons montré qu'en fonction du niveau d'abstraction de l'information visuelle, cette fluence

[137] : LITTLE et al. (2014), « Men's strategic preferences for femininity in female faces »

[138] : BANCHEFSKY et al. (2016), « But You Don't Look Like A Scientist! »

[139] : SCZESNY et al. (2006), « Masculine = Competent? »



pouvait être liée positivement ou négativement avec la beauté, même si prise plus globalement, elle explique la variance d'une part importante de la beauté perçue.

# Évaluer expérimentalement une préférence en écologie comportementale

# 9

En termes de mise en pratique technique, l'approche expérimentale est très liée aux contraintes concrètes du terrain dans laquelle elle est mise en œuvre. Nous reviendrons en 2.1 sur le dispositif expérimental utilisé, ses limites et comment nous les avons prises en compte. Ensuite, nous évoquerons en 2.2 la poursuite de ces travaux, en particulier des analyses complémentaires qui nous permettraient de renforcer nos résultats. Enfin, au-delà du dispositif spécifique que nous avons utilisé, nous discuterons en 2.3 des limites du paradigme du temps d'observation sur lequel nous sommes basés.

9.1 Regard critique sur le dispositif expérimental	113
9.2 Les limites du paradigme du temps d'observation	114
9.3 Analyse complémentaires et poursuite de ces travaux . . . . .	115

## 9.1 Regard critique sur le dispositif expérimental

Les mandrills étudiés sont issus d'une population en captivité. Bien que contrairement à d'autres primates du centre, les mandrills disposent d'espace de plusieurs hectares correspondant à l'écosystème des forêts tropicales des mandrills sauvages, leurs conditions de vie sont très différentes. De fait, des mandrills sauvages peuvent parcourir plusieurs dizaines de kilomètres par jour sans repasser au même endroit (observation personnelle). De plus, les individus en captivité sont nourris tous les jours, avec un régime principalement composé de fruits et de pain, tandis que les mandrills sauvages se nourrissent de ce qu'ils arrivent à trouver dans la nature, ce qui est très varié et peut comporter des racines, des insectes, des feuilles, des fruits et dans de rares cas de petits animaux comme des tortues (observations personnelles). Les populations de mandrills sauvages évoluent par reproduction, mais aussi par la migration des mâles à l'adolescence qui ne reviennent que très rarement dans le groupe dont ils sont originaires. Les mandrills captifs ne migrent pas, et en dehors de la reproduction, leur population peut parfois se voir augmentée d'individus qui ont été saisis chez des particuliers qui les détenaient illégalement. Enfin, la taille des groupes des mandrills captifs est particulièrement petite (environ 50 individus), ce qui peut parfois exister dans la nature même si les groupes sauvages sont plus grands la plupart du temps.

Ces différences entre des mandrills sauvages et la population captive étudiée ont probablement un impact sur la physiologie et les comportements sociaux-sexuels des mandrills. Les mâles captifs sont plus colorés, ce qui peut s'expliquer par une pression de sélection sexuelle plus importante du fait de l'absence de dispersion des mâles [127], et plus petits, ce qui peut s'expliquer par leur régime alimentaire particulier. Ces différences impliquent de relativiser les résultats obtenus avec cette population.

Néanmoins, des expériences comportementales avec des primates sont parfois réalisées avec des populations captives présentant encore plus de différences avec les populations sauvages. Par exemple, des travaux chez les chimpanzés ont utilisé un dispositif similaire au notre mais dans des conditions encore plus contrôlées ou les animaux avaient été habitués

[127] : SETCHELL et al. (2006), «Signal content of red facial coloration in female mandrills (*Mandrillus sphinx*)»

[115] : LEWIS et al. (2023), « Bonobos and chimpanzees remember familiar conspecifics for decades »

[128] : TIEO et al. (2023), « Social and sexual consequences of facial femininity in a non-human primate »

à se positionner face à des écrans sur de longues durées pour réaliser des séries d'expériences de ce type [115]. Dans la population que nous avons étudiée, les expérimentations comportementales sont beaucoup plus rares et les individus sont ainsi moins facilement biaisés par des souvenirs de dispositifs similaires qu'ils auraient eu. De plus, nous avons obtenu des résultats similaires à ceux trouvés sur une population de mandrills sauvages avec des types d'analyses différentes [128], ce qui permet d'augmenter leur crédibilité malgré ces biais.

Les chambres d'expérimentations que nous avons utilisées présentent aussi leurs limites. Premièrement, il s'agit de cages, qui n'isolent pas totalement les individus de leur environnement extérieur, ces individus étaient en effet visiblement perturbés par les autres mandrills avec qui ils ont interagi régulièrement pendant les essais. Les contraintes techniques liées à l'organisation du lieu dans lequel nous étions, ainsi que la tendance des mandrills à détruire par jeu tout objet inhabituel auquel ils sont confrontés, ne nous ont pas permis de résoudre ce problème dans le temps imparti. Deuxièmement, ces cages n'étaient pas symétriques et tandis qu'un des deux écrans était proche d'un côté de la cage où se situait la chambre de nourrissage, vide, l'autre écran était du côté de l'enclos ou d'autres mandrills étaient présents. Nous avons néanmoins contrôlé ces deux derniers biais en mélangeant aléatoirement la latéralité de l'écran représentant le mandrill le plus féminin. Troisièmement, la largeur des chambres d'expérimentations n'était pas tout à fait identique entre les 3 que nous avons utilisés. Nous avons pallié ce problème en marquant au sol le plus grand rectangle commun aux 3 chambres et en annotant le regard des mandrills uniquement lorsque ceux-ci se trouvaient à l'intérieur du rectangle.

## 9.2 Les limites du paradigme du temps d'observation

On a vu en partie 4.1.3 de l'introduction générale que le paradigme du temps d'observation permet d'interpréter des choix réalisés lors de tests expérimentaux en se basant sur le temps de regard respectifs des individus pour chacune des options qui leur sont proposées.

La première limite de cette approche, est qu'elle évalue bien un choix, et non pas une préférence. Ces deux termes ne sont pas nécessairement liés : une personne peut préférer le café au thé mais choisir un thé dans un restaurant car le café est plus cher. Pour éviter une utilisation inadéquate du terme "préférence", il a été proposé le terme de "biais visuel" pour désigner le stimulus qui est préféré [113] et éviter des interprétations erronées. Par exemple, des travaux ont montré que dans certains contextes, les sujets regardent plus longtemps un objet nouveau qu'un objet familier [140], et dans d'autres contextes, c'est l'inverse [141]. Ainsi, nous ne pouvons pas vraiment conclure sur le lien entre temps de regard et familiarité tel quel, mais des travaux complémentaires ont montré que le degré de familiarité [142] et le type de stimuli [143] étaient des facteurs qui pouvaient influencer les résultats. Cet exemple montre bien qu'une augmentation du temps d'observation peut être multifactorielle, et que ces facteurs sont difficiles à interpréter. Ainsi, la conclusion la plus parcimonieuse devant une différence de temps d'observation sans éléments supplémentaires

[113] : WINTERS et al. (2015), « Perspectives »

[140] : FANTZ (1964), « Visual Experience in Infants »

[141] : FUJITA (1987), « Species recognition by five macaque monkeys »

[142] : RICHMOND et al. (2007), « Interpreting visual preferences in the visual paired-comparison task »

[143] : PARK et al. (2010), « Roles of familiarity and novelty in visual preference judgments are segregated across object categories »

peut être que “*les sujets sont capables de faire la différence entre ces deux stimuli*” [113].

De fait, pour aller plus loin dans les conclusions, il est nécessaire d’ajouter d’autres indicateurs comportementaux à celui du temps d’observation. Pour autant, des travaux ont montré des résultats différents voire contradictoires selon que le paradigme utilisé soit le temps d’observation ou d’autres mesures [144-146]. Ainsi, il est difficile de trouver comment traduire ce biais visuel en comportement. C’est un argument de plus pour affirmer que des résultats basés sur le paradigme du temps d’observation doivent être corroborés avec d’autres analyses.

Enfin, des travaux basés sur le paradigme du temps d’observation ont échoué à être répliqués [147] ce qui suggère que des facteurs de confusion ont pu entrer en jeu sans être identifiés et contrôlés.

### 9.3 Analyse complémentaires et poursuite de ces travaux

Dans nos travaux, nous avons essayé de prendre en compte les risques liés à notre approche, afin d’avoir des résultats les plus robustes possibles. Par exemple, nous avons d’une part veillé à caractériser le dispositif avec des variables pouvant être des facteurs de confusion (comme la pièce où se déroule l’expérience) (Cf. Ch.2), et d’autre part, à créer des stimuli avec une méthode permettant de maîtriser ces facteurs quand ils concernent directement les stimuli (Cf. Ch.1). Comme nous l’avons montré, notre méthode a permis d’éditer des stimuli selon un axe de féminité tout en maîtrisant les autres sources de variation faciale.

Des analyses complémentaires sont toujours en cours et n’ont pas été présentées dans le corps de cette thèse. Ainsi, j’ai mentionné à plusieurs reprises que le paradigme du temps d’observation ne permettait pas à lui seul de conclure sur des préférences mais qu’il devait être mis au regard d’autres descripteurs comportementaux. De fait, dans nos analyses, nous avons encodé à partir des vidéos des essais d’autres types d’information. Il s’agit de deux types de comportements bien connus des mandrills et qui pourront nous aider à interpréter les résultats. Le hochement de tête (head bob) est un signe d’agressivité voire de menace. La grimace est un signe de bienveillance qui peut indiquer une marque de respect à un individu avec un rang plus élevé, ou une proposition de rapports sexuels dans un climat de confiance (“come-here face”). De plus, nous avons encodé un troisième type de comportement, le regard intense, qui n’est pas spécialement observé dans la nature mais qui témoignait d’un intérêt particulier pour les stimuli de la part des mandrills, sans qu’il soit possible de le classer comme un comportement affiliatif ou agressif. Les analyses statistiques nous permettant de déterminer si de tels comportements arrivent plus ou moins souvent chez des femelles moins féminines sont encore en cours. De fait, d’après les résultats plusieurs fois mentionnés sur l’impact de la féminité des visages sur les comportements sociaux sexuels des mandrills sauvages, les femelles les moins féminines sont plus approchées et reçoivent plus de copulation, mais aussi plus d’agression de la part des mâles. Ainsi, nous avons pour hypothèse que les 3 types de comportements encodés dans les vidéos des essais (hochement de tête

[113] : WINTERS et al. (2015), « Perspectives »

[144] : HOFSTADTER et al. (1996), « Response modality affects human infant delayed-response performance »

[145] : AHMED et al. (1998), « Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? »

[146] : CHARLES et al. (2009), « Object permanence and method of disappearance »

[147] : BOGARTZ et al. (1997), « Interpreting infant looking »

agressif, grimace affiliative et regard intense d'intérêt) seront tous plus associés à des femelles moins féminines qu'à des femelles plus féminines. L'agressivité des mâles envers des femelles qu'ils jugent plus attractives peut apparaître de prime abord contre-intuitive mais s'explique par un mécanisme de coercition sexuelle où les mâles infligent des blessures ou du stress aux femelles pour augmenter leur succès reproducteur [148].

[148] : BANIEL et al. (2017), « Male violence and sexual intimidation in a wild



FIG. 9.1 : Différents niveaux d'édition d'un même portrait de mandrill selon l'orientation de la tête (production personnelle)

Un autre aspect de nos travaux que nous n'avons pas encore mentionné est l'utilisation de notre méthode d'édition des portraits de mandrill pour modifier une variable d'orientation de la tête, nous permettant de créer la simulation d'un mouvement, tout en conservant les caractéristiques faciales de l'individu (autrement dit, une animation). Cela permet d'augmenter l'attractivité des stimuli en général, indépendamment de leur niveau de féminité (Cf. Fig. 9.1).

# L'IA en écologie et en évolution

Nous allons ici discuter certains aspects de l'intérêt de l'IA en écologie et en évolution, d'abord en se focalisant sur un problème technique qui nous semble central, l'encodage de données réelles dans un espace latent en 1.1, enfin sous le prisme particulier de ce que peut apporter l'IA générative à ces disciplines en 1.2. L'idée de cette partie n'est pas de lister l'ensemble des travaux d'écologie et d'évolution utilisant de l'IA en général, démarche bien trop large et hors sujet par rapport à nos travaux, mais bien de s'intéresser à certains axes qui m'ont semblé cruciaux dans nos travaux et dans leurs perspectives.

<b>10.1 Encoder des données réelle dans un espace latent . . . . .</b>	<b>117</b>
<b>10.2 L'IA générative en écologie et évolution</b>	<b>119</b>
10.2.1 Augmenter et compléter des données . . .	119
10.2.2 Créer des stimuli . .	120
10.2.3 Simuler l'évolution .	120
10.2.4 Prédire le vivant . . .	121

## 10.1 Encoder des données réelle dans un espace latent

En sciences des données, il est traditionnel, avec des données de grandes dimensions, d'utiliser des projections de ces données dans des espaces de plus petites dimensions en essayant de perdre le moins de variance possible. Par exemple, l'analyse en composantes principales (ACP) permet de transformer des variables en d'autres variables, combinaisons linéaires des premières, décorréélées entre elles ce qui permet d'expliquer un maximum de variance dans les données avec moins de variables. Une analogie peut être faite avec des réseaux de neurones qui traitent une image : les premières variables sont les pixels, les données transformées sont les valeurs du vecteur latent, coordonnées de l'image dans l'espace latent. Dans le cas d'un CNN qui classe une image, l'image est donc directement projetée dans l'espace latent. Avec certaines architectures d'IA générative, nous observons la même approche : un encodeur va projeter les données dans un espace latent, afin de se servir de cet espace latent pour générer de nouvelles données similaires. Par exemple, avec un auto-encodeur, les images sont projetées dans un espace latent avec l'encodeur, et d'autres sont générées à partir du même espace latent avec le décodeur. En étudiant la géométrie des espaces latents, et les positions respectives des données et des variables, des informations utiles peuvent être obtenues : la proximité ou la distance entre des variables, ou encore leur classe.

Avec un GAN, c'est un peu plus compliqué car il n'y a pas d'encodeur. Le générateur fabrique des images en essayant de les faire correspondre à la distribution des caractéristiques d'images réelles, ce qui est évalué par le discriminateur grâce au mécanisme adversarial (voir partie 3.1.3 de l'introduction). Des travaux ont conçu des encodeurs permettant a posteriori de projeter des images réelles dans l'espace latent du générateur, qui prend alors un rôle de décodeur (Cf. encadrés 4 et 5). Néanmoins, les images réelles une fois décodées par le générateur peuvent être légèrement modifiées et altérées, car leurs caractéristiques ne correspondent pas strictement à la distribution apprise. Pour des visages de mandrills, l'identité même des individus semble être compromise par ce processus (Berta Roura Torres, communication personnelle). Pour des visages d'humains, c'est aussi le cas, et la distribution des caractéristiques des images

d'apprentissages est d'autant plus cruciale. Par exemple, en Fig.10.1, une image réelle du visage du doctorant (à gauche) est encodée avec l'encodeur pSp [pSp], puis décodée avec StyleGAN3 (à droite). Les deux images sont certes similaires mais loin d'être identiques. Tout d'abord, la forme et la texture du visage sont légèrement modifiées. Ensuite, le fond a complètement changé. C'est assez logique car le GAN entraîné à réaliser cette tâche est spécialisé dans des images de visages, mais pas dans les arrière plans qui peuvent être diverses et variés. Enfin, les lunettes sont complètement différentes. Pour comprendre pourquoi, on rappelle que les GANs sont entraînés à apprendre et à reproduire les caractéristiques d'une base de données réelles. La base de données sur laquelle le GAN a été entraîné, FFHQ [83], date du début des années 2000. La photo date elle de 2019. Le type de lunettes communément portées n'étant pas le même entre ces deux périodes (observation personnelle), le GAN les a donc "remplacées" par un style de lunettes qu'il connaît et qui existait dans sa base de données d'intérêt. Il s'agit d'un cas de surapprentissage : le GAN n'as pas réussi à généraliser à partir de sa base de données d'apprentissage sur des données plus générales.

[83] : KARRAS et al. (2019), *A Style-Based Generator Architecture for Generative Adversarial Networks*



**FIG. 10.1 :** Encodage d'une image réelle (à gauche) dans l'espace latent de StyleGAN3 entraîné sur une base de données de portraits d'humains (au milieu), et avec une méthode basée sur des modèles de diffusion (à droite) (production personnelle)

[84] : BERMANO et al. (2022), « State-of-the-Art in the Architecture, Methods and Applications of StyleGAN »

[122] : RICHARDSON et al. (2021), *Encoding in Style*

[87] : SOHL-DICKSTEIN et al. (2015), *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*

[149] : BRACK et al. (2024), *LEDITS++*

La force de StyleGAN3, au delà de son réalisme est qu'il est capable de rendre linéaires dans son espace latent les variables qui décrivent le plus la variance des images (comme le sexe pour un visage)[84, 122]. De plus, quand nous avons commencé ces travaux en 2021, il s'agissait de l'état de l'art de la génération d'images synthétiques. Depuis, des technologies très prometteuses ont émergé, comme les modèles de diffusion [87](qui existaient avant mais qui se sont considérablement améliorés récemment, cf. partie 3.1.3 de l'introduction). Nous n'avons pas étudié en quoi ce type d'architecture pourrait être une approche complémentaire ou meilleure que celle basée sur les GANs. Toutefois, à titre exploratoire, nous avons utilisé une interface graphique permettant d'encoder des images réelles avec un modèle de diffusion [149] qui, visuellement, donne des résultats beaucoup plus performants que l'encodeur lié à StyleGAN3 (Cf. Fig. 10.1). De fait, cela semble une voie d'amélioration prometteuse pour nos travaux, dans un contexte scientifique où les technologies d'intelligence artificielle générative s'améliorent très rapidement.

## 10.2 L'IA générative en écologie et évolution

Dans cette partie, nous allons nous éloigner des questions de recherches propres à cette thèse pour nous intéresser à ce que peut apporter l'IA générative à des domaines variés de l'écologie et de l'évolution. Il s'agit des premiers pas d'une réflexion au long cours ayant pour finalité l'écriture d'un article de revue de la littérature de ce champ. Nous avons réuni par thèmes les solutions qu'apporte l'IA générative à nos disciplines, et nous développerons comment les chercheurs ont pu trouver des solutions à des grandes thématiques.

### 10.2.1 Augmenter et compléter des données

Dans les disciplines de ces travaux, comme dans d'autres, la question des données est fondamentale. Prélever des données réelles sur le terrain peut être long, fastidieux et coûteux. Pourtant, pour comprendre la manière dont ces données se structurent et ainsi répondre à des questions de recherche, il est parfois nécessaire d'en avoir beaucoup. L'IA générative offre des solutions pour augmenter artificiellement des données. Même s'il existe des techniques plus traditionnelles pour faire cela, l'avantage de l'IA générative est son réalisme dans la mesure où elle est capable de modéliser finement les caractéristiques des données.

Plusieurs exemples illustrent cela. Pour entraîner des algorithmes de classification d'image, il est préférable que les différentes classes de la base de données d'entraînement soient équilibrées. Parfois, ce n'est pas le cas, et c'est difficile de trouver de nouvelles données pour les rééquilibrer. Les techniques traditionnelles d'augmentation de données consistent à dupliquer les données en les déformant ou en les tournant. Des travaux ont montré qu'avec des images d'insectes parasites de plusieurs espèces, il est possible d'entraîner des modèles de classification plus performants quand la base de données d'entraînement avait été rééquilibrée (en complétant les classes déficitaires) avec des GANs qu'avec les techniques traditionnelles [150]. Des résultats similaires ont été obtenus dans un objectif de classification de cartes de zones humides, avec des données différentes, en 3 dimensions, et un autre type d'architecture de GAN [151], et pour classer des plantes [152]. Enfin, des travaux en épidémiologie évolutive ont eu une approche similaire pour augmenter des données dans le but de prédire l'évolution de la pandémie de Covid-19 [153]. L'originalité de leur approche consiste à ne pas se servir de données réelles pour entraîner leurs GANs mais plutôt les résultats de modèles épidémiologiques traditionnels basés sur la subdivision entre les individus susceptibles, infectés et guéris, traditionnellement utilisés dans ce champ disciplinaire [154].

Pour autant, cette approche a ses limites. En effet, les modèles d'IA générative reproduisent la distribution des caractéristiques des bases de données mais ne vont pas créer des nouvelles caractéristiques, simplement les "mélanger". Il n'y aura donc pas d'information nouvelle créée avec ces approches. Dit autrement, l'IA générative ne sait pas extrapoler, elle ne fait que interpoler à l'intérieur de son domaine d'apprentissage [155]. Nous avons pu nous en rendre compte en montrant des mandrills synthétiques à une experte habituée à les reconnaître individuellement sur le terrain, qui nous a dit que certains de ces faux mandrills lui semblaient

[150] : LU et al. (2019), « Generative Adversarial Network Based Image Augmentation for Insect Pest Classification Enhancement »

[151] : JAMALI et al. (2022), « 3DUNetGFormer »

[152] : MADSEN et al. (2019), « Generating artificial images of plant seedlings using generative adversarial networks »

[153] : WANG et al. (2023), « Four-channel generative adversarial networks can predict the distribution of reef-associated fish in the South and East China Seas »

[154] : BENHAMOU et al. (2023), « Phenotypic evolution of SARS-CoV-2 »

[155] : HASSON et al. (2020), « Direct Fit to Nature »



être des mélanges entre deux ou trois mandrills réels (Berta Roura Torres, communication personnelle).

### 10.2.2 Créer des stimuli

Dans nos travaux, nous avons utilisé des GANs pour fabriquer des stimuli pour des expériences en écologie visuelle, démarche dont nous avons expliqué l'intérêt dans la partie 4 de l'introduction. D'autres travaux ont suivi une démarche similaire, mais avec des différences de technologies et de modèles d'études.

Ainsi, le transfert de style, une tâche d'apprentissage profond permettant de transférer le style d'une image sur le contenu d'une autre image, a été utilisé<sup>1</sup>. L'idée était de transférer le style de certains habitats de poissons sur la peau de ces poissons pour tester si certains poissons étaient préférés dans des tests de choix et ainsi tester une hypothèse postulant que les motifs des poissons avaient évolué pour correspondre aux caractéristiques spatiales de leur environnement [156].

1 : un exemple connu de cette tâche est la transformations d'images représentant toute sortes de choses avec des styles caractéristiques de certains peintres, comme Van Gogh

[156] : HÉJJA-BRICHARD et al. (2023), « Using generative artificial intelligence to test hypotheses about animal signal evolution »

[157] : TALAS et al. (2020), « CamoGAN »

### 10.2.3 Simuler l'évolution

L'apprentissage automatique et l'évolution des espèces présentent des points communs. L'optimisation des paramètres des modèles d'apprentissage automatique peut être comparée à la sélection des mutations génétiques [158]. Alors que certaines approches d'apprentissage auto-

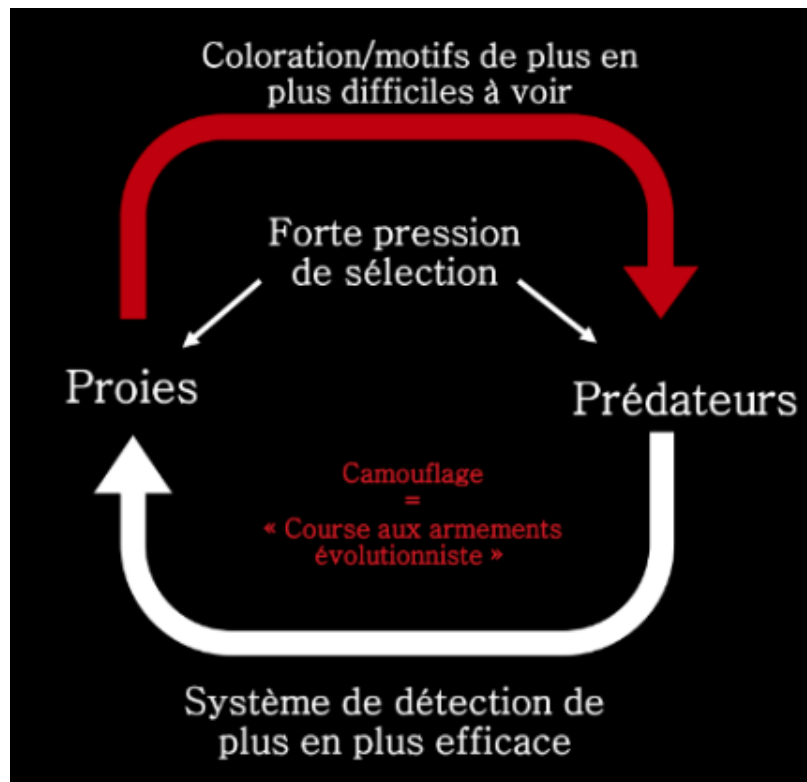


FIG. 10.2 : co-évolution entre le système visuel des prédateurs et le phénotype de camouflage des proies (production personnelle, illustrant les travaux de [157])

matique, comme l'apprentissage par renforcement ou les algorithmes génétiques comportent encore plus d'homologies avec l'évolution, l'IA générative peut aussi y être assimilée et même servir de modèle pour simuler

l'évolution. Des travaux ont proposé que le mécanisme adversarial des GANs pouvait être comparé à la relation proie-prédateur, précisément les vecteurs latents du générateur au génotype des proies et le discriminateur au système visuel des prédateurs. De fait, les deux sont soumis à une forte pression de sélection : les prédateurs pour avoir des systèmes de détection les plus efficaces, et les proies pour avoir des couleurs et des motifs de plus en plus difficiles à voir (Cf. Fig. 10.2). Ils ont ainsi modélisé avec un GAN la co-évolution entre la détection de papillons sur des arrières plans et le camouflage de ces papillons pour ressembler à l'arrière-plan [157].

[158] : ARAK et al. (1997), « Hidden preferences and the evolution of signals »

Cette approche, très concrète, n'est pas la seule. D'autres travaux ont eu une approche plus théorique en développant l'idée que le mécanisme des autoencodeurs pouvait être assimilé à la structure des êtres vivants. Pour les auteurs de ces travaux, l'élément fondamental du vivant qui évolue n'est ni l'individu, ni le gène mais les interactions entre les différents "codes" des êtres vivants : génome, protéome, etc. Cette approche, certes séduisante si elle est considérée comme une abstraction intellectuelle, semble néanmoins peu réaliste et ne fait pas du tout l'objet de consensus [159].

[157] : TALAS et al. (2020), « CamoGAN »

#### 10.2.4 Prédire le vivant

L'IA générative peut aussi être utilisée dans un but de prédiction, que ce soit dans une dimension spatiale ou temporelle. Des travaux ont utilisé un GAN pour prédire des co-occurrences d'espèces végétales, c'est-à-dire quelles plantes avaient plus de probabilité de se trouver proche d'autre type de plantes [160], à partir de vecteurs de présence-absence de ces plantes. Les modèles ainsi entraînés peuvent servir de base pour réaliser de l'apprentissage par transfert avec d'autres environnements. Des données plus complexes peuvent aussi bénéficier des capacités prédictives des GANs dans un contexte temporel. Ainsi, des travaux ont réalisé des prédictions de distributions de poissons récifaux en mer de Chine sur des variables saisonnières de probabilité de présence de différentes espèces, avec une architecture de GAN particulière : le retro-cycle-GAN. Il s'agit d'une architecture qui permet d'utiliser des données générées en tant que données d'apprentissage de manière cyclique, permettant des modèles plus robustes en particulier dans un contexte de prédiction de séries temporelles [153]. Enfin, des GANs ont été utilisés en écologie du déplacement pour prédire les trajectoires d'oiseaux marins [161]. Ces travaux sont d'ailleurs poursuivis pour comparer aux GANs d'autres architectures comme des modèles de diffusion sur la même tâche.

[159] : COHEN et al. (2023), « Evolution is driven by natural autoencoding »

[160] : HIRN et al. (2022), « A deep Generative Artificial Intelligence system to predict species coexistence patterns »

[153] : WANG et al. (2023), « Four-channel generative adversarial networks can predict the distribution of reef-associated fish in the South and East China Seas »



Dans ces travaux, nous nous sommes intéressés aux mécanismes de l'attractivité des visages, chez deux espèces, sous les angles différents et complémentaires de la modélisation et de l'expérience empirique. Ces approches nous ont permis de répondre à des hypothèses sur l'attractivité des visages, à travers l'un de ses déterminants, la féminité, mais aussi plus globalement sur la manière de modéliser la beauté.

Au-delà des réponses aux questions scientifiques, nous avons proposé des approches basées sur des développements récents en intelligence artificielle, plus particulièrement avec des réseaux de neurones, génératifs et prédictifs.

Ces travaux peuvent servir de bases à des travaux futurs. Notre approche de modélisation de la fluence propose de modéliser l'expérience esthétique en s'intéressant à la manière dont est traitée l'information dans le système visuel : ici à travers la sparsité du codage neuronal modélisée par un réseau de neurones artificiels. Ce paradigme pourra être utilisé dans des études futures, explorant par exemple d'autres formalisations mathématiques de la fluence dans de tels réseaux. Utiliser l'IA générative pour créer des stimuli pourra non seulement permettre des travaux dans le domaine de la communication visuelle des primates, mais aussi en psychologie humaine. Plus généralement, l'IA générative semble être un outil très prometteur pour tester des hypothèses et apporter des éléments de réponse en écologie et en évolution.



# **APPENDIX**



# Comparing activation typicality and sparsity in a deep CNN to predict facial beauty

Sonia Tiew, Melvin Bardin, Roland Bertin-Johannet, Nicolas Dibot, Tamra Mendelson, William Puech\* Julien P. Renoult\*

\*Equal contribution

## **ABSTRACT**

Processing fluency, which describes the subjective sensation of ease with which information is treated by the sensory systems and the brain, has become one of the most popular explanation to aesthetic appreciation and beauty. Two metrics have recently been proposed to model fluency: the sparsity of neuronal activation, which characterizes the efficiency of neural processing, and the statistical typicality, which describes how well the encoding of a stimulus matches a reference representation of stimuli of the same type. Using Convolutional Neural Networks (CNNs) as a model for human perception, this study compares the ability of these metrics to explain variation in facial attractiveness. Our findings show that the sparsity of neuronal activations is a more robust predictor of facial beauty than statistical typicality. Refining the reference representation to a single ethnicity or gender did not increase the explanatory power of statistical typicality. Overall, our results highlight the significance of neural processing efficiency in aesthetic judgments of facial beauty.

**Key-words:**



## INTRODUCTION

Beauty holds significant influence across multiple aspects of human life. It shapes our perceptions, judgments (known as the "halo effect" [1]), preferences [2], and thereby guides our decision-making across diverse arenas, such as personal relationships [3] and consumer choices [4]. In the realm of cognitive science, processing fluency—defined as the sensation ease of interpreting sensory information—has gained significant attention as a plausible determinant of people's evaluation of beauty [5]. While fluency is currently subjectively measured through psychological experiments, few studies have attempted to model it, and none have compared the capabilities of existing models to predict beauty.

As a concept in aesthetic psychology, processing fluency provides a powerful explanation for a wide range of aesthetic inclinations, for both simple and complex stimuli. A preference for basic features such as symmetrical visual patterns and high contrasts is explicable by their effortless processing [5-7]. More complex stimuli, like fractal patterns, would be appreciated for their smooth processing as well, since the self-similarity of fractals at different scales makes these patterns highly predictive [8,9]. Fluency would also explain the attractiveness of prototypical representations [10]. Prototypes are typified by familiar and easily discernible features, and thus prototype-like stimuli are processed with ease, enabling rapid and accurate categorization while enhancing memory retention. Regardless of their specific attributes, prototypes consistently earn preference across an array of stimuli, encompassing biological, inanimate, and abstract forms [10-12].

Fluency, therefore, has a high explanatory power, and using this concept to predict beauty would have numerous technological applications, but also fundamental implications, for instance, in allowing to study beauty in non-human animals [13]. However, the explanatory and predictive capacity of fluency is limited by the scarcity and current limits of studies aiming to model this concept. Early research in modeling processing fluency primarily focused on objective measures of feature repetitions in stimuli, such as symmetry, contrast, and self-similarity, for visual stimuli [14]. While these metrics could provide valuable insights, they predominantly focus on the stimulus itself,

assuming that the studied features ease information processing. Yet fluency is fundamentally rooted in the interaction between the stimulus and the perception of the beholder, and thus more accurate metrics should target features as they are processed by the perceptual system.

A first step in modeling fluency is therefore to model the processing of features using a model of perception. The development of such models is uneven across the different sensory modalities, and is by far the most advanced for visual perception, on which we will focus in this study [13,15]. As demonstrated in numerous studies [16,17], Deep Convolutional Neural Networks CNNs have recently emerged as powerful models of visual information processing, from low-level feature extraction to high-level semantic interpretation. Just like how our visual system processes information, CNNs start by extracting simple features such as edges and contours in their initial layers. As the information flows through the network, the recognition of increasingly intricate features takes place, which parallels our visual system's ability to discern complex objects or scenes by combining simpler constituent elements. By studying neuronal activations within CNNs, which represent the response of different neurons of each layer to specific image inputs, we can thus gain valuable insights into how biological vision processes visual information.

The second step in modeling fluency is to choose a metric that characterizes the ease of information processing. Inspired by the information theory applied to biological systems [18], some authors have proposed to model fluency using the sparsity of the neuronal activation [14,19]. Sparsity measures the concentration of neuronal activity in a specific subset of features or patterns. A sparse stimulus thus activates only a few neurons simultaneously, leading to a low-cost, efficient processing of information [20]. Using the sparsity of neuronal activations to estimate fluency fits the prediction of Winkielman et al. [21], that “fluent patterns should be represented by more extreme values of activation”. Previous studies have provided empirical evidence supporting the link between sparsity and beauty. For instance, using a model of information processing in the primary visual cortex, [19] have shown a positive correlation between the sparsity of neuronal activations and the perceived attractiveness

of female faces. Furthermore, sparsity has been identified as a robust predictor of face attractiveness compared to other factors such as body mass index, sexual dimorphism, averageness, and asymmetry [22]. Very recently, one study evaluated the ability of the sparsity of neuronal activations within a CNN to explain variation in the beauty of faces and artistic paintings [23]. The authors showed that sparsity alone could explain up to 28% of the variance in beauty scores.

Another metric of fluency proposed in the literature is the statistical typicality. Typicality describes the extent to which a stimulus aligns with an average representation. It is based on the underlying assumption that individuals form mental representations of averageness for various categories based on their past experiences and exposure to stimuli. Typicality is thus closely related to familiarity, which has been a well-studied factor influencing fluency in psychological studies [7]. Ryali and colleagues [24] demonstrated that the attractiveness of a face is partly explained by its statistical typicality, defined as the likelihood of the face image relative to an internal representation of the face distribution. The authors used the Active Appearance Model (AAM) as a model of face perception, which is built from features describing the shape and texture of faces. They then showed that the attractiveness of a given face can be predicted from its likelihood estimated from the distribution of faces of the same gender.

The contribution of this study is threefold. First, we propose a new model of fluency that applies the statistical typicality metric to convolutional neural networks (CNNs). More precisely, for each layer of the CNN the method estimates the likelihood of a stimulus encoding given a reference distribution of encoded features. The method thus extends previous applications of the statistical typicality ([24]; see also [25] ) to a model of perception that describes visual information processing as it operates throughout the retina and the ventral stream of the visual system, and that is not specific to one visual domain (as is, e.g., the AMM model). Second, we aim to compare the ability of sparsity and statistical typicality computed in CNN layers to explain variation in the attractiveness of human faces. Attractiveness is arguably the strongest determinant of facial beauty, to a point that both terms are

generally used interchangeably in the psychological literature [2]. For this purpose, we used the publicly available Chicago Face Dataset [26], a comprehensive collection of face images with empirical scores of attractiveness. We encoded each face with a CNN, calculated its sparsity and likelihood at each layer, and trained two regression models (one with sparsity, the other with likelihood) to predict attractiveness. Third, we examine the extent to which the ability of statistical typicality to explain facial attractiveness is influenced by the choice of the reference distribution. Specifically, we investigate whether specializing the reference distribution to include only faces of a single gender and/or a single ethnic group improves the ability of likelihood to predict attractiveness.

## **MATERIAL AND METHODS**

### **1. Materials**

We modeled the fluency of processing portrait images of Chicago Face Dataset (CFD hereafter; [26]). This dataset contains standardized photographic portraits of individuals aged 17 to 65, spanning a variety of ethnic backgrounds, including East Asian, Black, Hispanic, and White, with balanced representation across genders. We use a subset of 597 portraits of CFD, each depicting a neutral expression, frontal view, and standardized attire. This dataset also includes ratings of attractiveness, evaluated by independent judges. Each portrait is associated with one mean score of attractiveness.

### **2. Modeling face processing using CNNs**

To model the visual processing of faces, we compared two Convolutional Neural Networks (CNNs), VGG16 [27] and VGGFace [28], that have been pre-trained on the ImageNet and VGGFace datasets, respectively. VGG-16 includes 13 convolutions and two fully connected layers. ImageNet is a large dataset of 14 million images depicting about 20,000 categories including people, plants, animals and human-made objects. VGG16 trained on such a large and varied dataset allows modeling a visual cortex that is not specialized

to one specific task [29,30]. In contrast, VGGFace is a variant of VGG16 specifically tuned for face recognition. The VGGFace dataset includes images capturing variations in facial expressions, angles, and lighting conditions. The fine-tuning of VGGFace to this dataset allows the entire network to adapt to face-specific features across all layers.

### **3. Metrics of fluency**

In a CNN, the output of a convolutional layer is a matrix of size  $H \times W \times C$  where each entry represents the activation of a “neuron”. The dimension  $C$  describes the number of channels, each embodying a unique feature, such as a distinct contrast or edge orientation.  $H \times W$  is termed a feature map, describing where in the image a feature is present;  $H$  and  $W$  are the height and the width of the feature map, respectively.

#### **Statistical typicality**

The first metric of fluency investigated is statistical typicality, calculated in three steps. The first step reduces the dimensionality of the  $H \times W \times C$  matrices. As the images traverse through a pre-trained network, a multitude of feature maps are generated from the different layers. Some of these feature maps have an extensive number of activations (for instance, the first convolution layer of VGG16 outputs 3,211,264 activations), rendering the estimation of statistical typicality computationally intractable. We compared three strategies of dimensionality reduction. In the first strategy, we performed one Principal Component Analysis (PCA) per layer (“layer-wise PCA”), thus considering all activations (ie, after flattening the  $H \times W \times C$  matrix). We kept the number of principal components allowing to account for 80% of the variance in activations [32]. This number varied across layers and datasets (between 28 and 622). In the second strategy, a PCA was applied individually to each feature map, preserving the components allowing to account for 80% of intra-map variance. The components from all feature maps within a specific layer were then concatenated and reduced further using a second PCA, again keeping the number of principal components allowing to account for 80% of the variance. In the third strategy, one first calculated the mean activation of each feature map and then concatenated all the means into a single vector for

each layer. Despite their potential interest, in particular the low computation cost of the third strategy, neither the second nor the third strategy improved our results compared to the first one. We thus here only present results obtained with the first strategy, based on a layer-wise PCA.

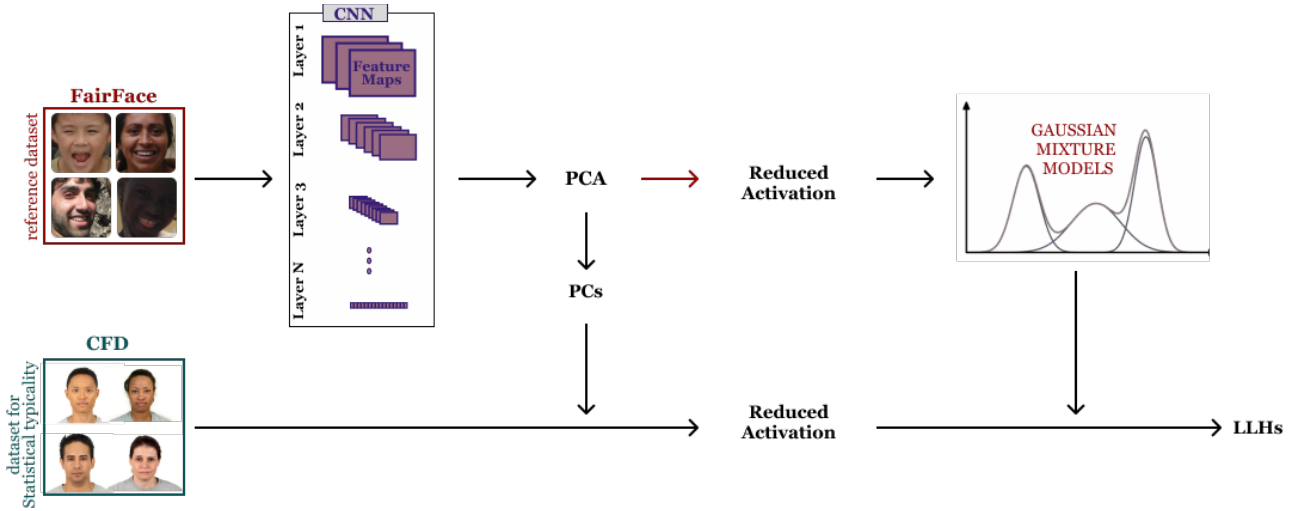
The second step for measuring statistical typicality was to establish a reference distribution of encoded features. Following the approach proposed by others [24,25], reference distributions were represented by Probability Density Functions (PDFs; one PDF per layer) from the reduced (by the layer-wise PCA) activations of all encoded portraits of a reference dataset. We used FairFace as a reference dataset [33], a collection of face portraits with a perfectly balanced representation of ethnic, gender, and age classes, enabling us to build an unbiased model of statistical typicality. PDFs were estimated from a balanced subset of 1,000 portrait images from FairFace. This number was chosen to allow a comparison of results when changing the reference dataset (see the section on the *Influence of the reference dataset*).

We estimated each PDF with Gaussian Mixture Models (GMMs), using the principal components as the variables in the Gaussian models. With their capacity to combine simple Gaussian density functions or "Gaussian components", GMMs adeptly capture the complexity of the underlying distribution of facial features. We determined the optimal number of Gaussian components with the Bayesian Information Criterion (BIC), setting an upper limit to 20 components.

In the third step, we used PDFs calculated from the reference datasets to estimate the log-likelihood (LHH) of each image of the Chicago Face Database (CFD), for every layer of the CNN. This was done after projecting the activation values of each image of CFD onto the principal components calculated using the FairFace dataset.

One limitation of GMMs is their sensitivity to singularities. Concretely, this means that one gaussian component can be fitted onto a single outlier, thus inflating the LLH of images located close to this outlier in the feature space. To mitigate this potential issue, the LLH of an image was calculated as the median value obtained after 100 replications of the analysis (including the

second and third steps described previously, but not the first, dimensionality reduction step, because the layer-wise PCA is deterministic and thus was performed only once). Using Bayesian Mixture Gaussian Models as an alternative to address the problem of singularities yielded similar results but at much higher computational cost (results not shown).



**Figure 1. Pipeline for estimating the statistical typicality of faces.** Images of the FairFace dataset are first encoded with a CNN (VGG16 or VGGFace). For each layer of the CNN, the dimensionality of the activation space describing the encodings is then reduced using a layer-wise principal component analysis (PCA), keeping only the principal components (PCs) explaining 80% of the variance. One probability density function (PDF) per layer is estimated using Gaussian Mixture Models (GMM) fitted onto the retained PCs. In a second step, images of the CFD dataset are encoded as in the first step, and their log-likelihood (LLH) calculated for each layer using the previously calculated PDFs. This entire process is repeated 100 times, and the statistical typicality is eventually calculated as the median of all repetitions.

## Sparsity

The second metric of fluency is activation sparsity, which measures the concentration of neuronal activity in specific features. We used the method described in detail in [23]. Briefly, we quantified sparsity using the Gini index [31]. To do so, we flattened the activation matrices into one-dimensional vectors and sorted the vectors in ascending order. The Gini index was then computed using the formula:

$$Gini = (\sum(2 * i - n - 1) * x_i) / (n * \sum x_i)$$

where,  $n$  is the total number of activations in the layer (vector length) and  $x_i$  the activation value at index  $i$ . Higher Gini indices indicate a higher degree of sparsity, reflecting a more selective activation.

#### 4. Statistical analyses

Using the face portraits of CFD, we assessed the ability of layer-wise statistical typicality (LLHs) and sparsity of activations calculated from all convolutional and fully connected layers ('AllLayers' models, see below) to predict facial attractiveness. We conducted a regression analysis with attractiveness as the response variable, and the LLHs or sparsity values of every layer as independent variables. Our first model involving statistical typicality can be expressed as:

$$\text{Attractiveness} \sim \sum(a_l * \text{LLH}_{\text{layer}_l}),$$

with  $l$  varying from 1 to  $N$ ,  $\text{LLH}_{\text{layer}_l}$  representing each layer's *LLH* and  $a_l$  the regression coefficient.

We similarly defined our model for sparsity as:

$$\text{Attractiveness} \sim \sum(b_l * \text{Sparsity}_{\text{layer}_l}),$$

with  $l$  varying from 1 to  $N$  (the number of layers for VGG or VGG16),  $\text{Sparsity}_{\text{layer}_l}$  indicating the sparsity of layer  $l$  and  $b_l$  the regression coefficient.

We employed ridge regression models, rather than classical linear regression models, to address the inherent collinearity among layers, due to the fact that layer outputs are inputs of the following layers. Ridge regression effectively handles multicollinearity and overfitting by applying a regularization that balances model complexity and generalization [34]. We performed ridge regressions using the *glmnet* package in R and a 10-fold cross-validation repeated 10 times. For ridge regression ( $\alpha=1$ ), we explored a range of penalty values ( $\lambda$ ), spanning from  $10^2$  to  $10^{-4}$ . Both the predictors and the



response were centered and scaled prior to the analysis. The coefficient of determination,  $R^2$ , is the explained variance of attractiveness.

In addition to our primary focus on 'AllLayers', we delved into the distinct subsets of the VGGFace neural network layers. These subsets included 'LastConvLayer', 'PoolingLayers', 'ShallowLayers', 'MiddleLayers', and 'DeepLayers'. Each subset represents a specific arrangement of layers, with breakdowns detailed in the supplementary information Figure S1. By analyzing these subsets, we aimed to unravel how different layers of the neural network contribute to facial attractiveness predictions.

## **5. Influence of the reference dataset**

Previous research suggested that the categorization of faces according to gender and ethnicity can impact perceived attractiveness [35,36]. We then investigated if a more specialized reference Probability Density Function (PDF) — built using a reference dataset tailored more specifically to a particular gender, ethnic group, or a combination of the two — could enhance the ability of statistical typicality to explain variation in attractiveness.

We used the gender and ethnic categories of FairFace to build 15 reference datasets, each containing exactly 1,000 images (see details in Table 1). These 15 datasets can be categorized into four types of reference datasets differing in the level of specialization: “all”, “ethnicity”, “gender” and “ethnicity x gender”. We consider that “ethnicity x gender” represents a more specialized reference dataset than “ethnicity” and “gender”, which are themselves more specialized than “all”. Ethnic and gender categories were kept balanced within the “all”, “ethnicity” and “gender” reference datasets. As previously, we estimated PDFs (one per layer) for each reference dataset. Then, for each image of one combination of gender and ethnic category of CFD, we calculated its LLHs with the different reference datasets. For Asian male individuals, for example, we calculated LLHs of all images of Asian males of the CFD dataset considering PDFs estimated from a reference dataset of 1,000 portraits of FairFace depicting either i) individuals of both genders and

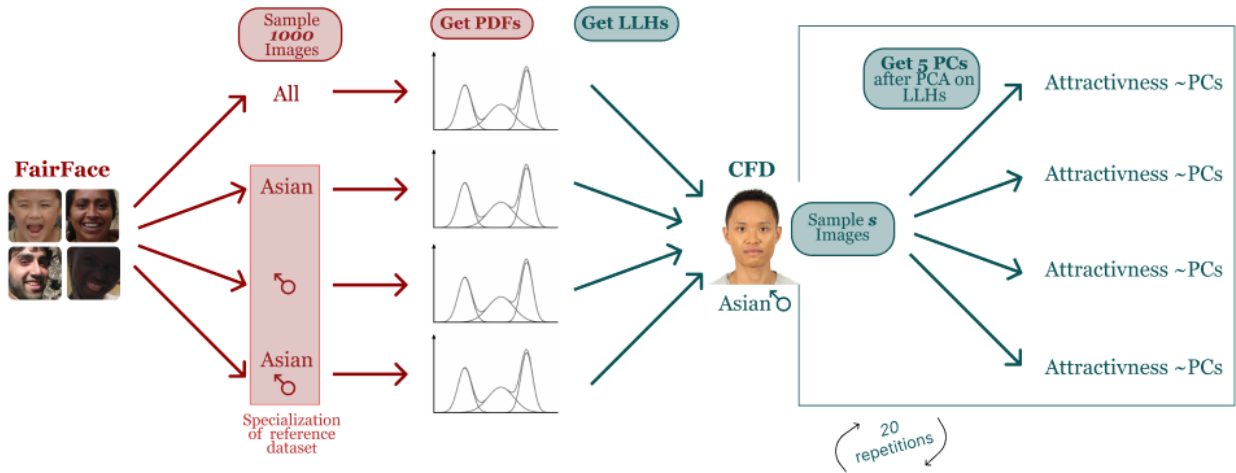
all (Asian, Black, White, Latino) ethnic groups (“all” reference dataset), ii) all Asian males and females (“ethnicity”), iii) males of all ethnic groups (“gender”), and iv) Asian males only (“ethnicity x gender”) (Figure 2).

To mitigate the problem of having a different number of images between gender and ethnic categories in CFD ( $R^2$  is influenced by the number of observations, and adjusted- $R^2$  cannot be calculated in a ridge regression), we randomly sampled 52 images from each category, equivalent to the minimum number of images present in any category. This sampling was repeated 20 times.

For each reference dataset, we thus obtained a set of LLHs (one per layer), which were as previously used in a ridge regression model to explain facial attractiveness. However, to address the challenge of having a large number of regressors relative to the number of observations (only 52 observations), rather than using raw LLHs values in the regression models, we reduced the number using a PCA. The first 5 principal components (PCs) cumulatively explained between 92% to 98% of variance. We calculated  $R^2$  for each ridge regressions, using cross-validation as above. To test the hypothesis that a more specialized reference dataset leads to higher  $R^2$  values, we then used a Generalized Linear Mixed Model (GLMM) with the  $R^2$  (derived from the mean of the 20 repetitions) as the response variable and the type of reference specialization (categorical, four levels, see Table 1) as explanatory variable. The categories of “ethnicity x gender” within the CFD dataset (CFD\_Categories\_“ethnicity x gender”) were included as a random effect (categorical variable, eight levels, see Table 1) in the model. The model was thus expressed as:

$$R^2 \sim PC1(LLH) + PC2(LLH) + PC3(LLH) + PC4(LLH) + PC5(LLH) + ReferenceSpecializationType + 1|CFD\_Categories\_“ethnicity \times gender”$$

The model was fitted using lme4, and results presented using sjPlot in R.



**Figure 2. Log-Likelihood (LLH) calculation using different reference datasets - An example using the Asian males category.** LLH values for Asian males in the Chicago Dataset were calculated using four types of reference dataset: 'all', 'Asian' (ethnicity-specific subset from FairFace), 'Males' (gender-specific subset from FairFace), and 'Asian Males' (ethnicity and gender-specific subset from FairFace). To address the variation in the number of images for each gender and ethnic category within the CFD dataset (see Table 1), we randomly selected a set of  $s=52$  images from each category, which corresponds to the smallest category size. The derived LLH values underwent dimensionality reduction using PCA, resulting in the first 5 principal components (PCs). These 5 PCs were then incorporated into a ridge regression model trained to predict facial attractiveness. This sampling process was carried out 20 times to ensure consistency. The process delineated in the figure is representative of the method applied to all 15 categories (Table 1).

**Table 1. Description of Reference datasets built from FairFace dataset.** Each reference dataset includes 1,000 images randomly sampled from the entire FairFace dataset (FairFace\_All), or from a subset of a single gender, a single ethnicity or a single gender of a single ethnicity. The column "Genders & ethnic groups" indicates the composition of the subset.

Genders & ethnic groups	Reference specialization type	Reference dataset identifier
all	all	FairFace_All
Asian, both genders	ethnicity	FairFace_A
Black, both genders	ethnicity	FairFace_B
Latino, both genders	ethnicity	FairFace_L
White, both genders	ethnicity	FairFace_W
Females, ethnicity	gender	FairFace_F

males, all ethnicity	gender	FairFace_M
Asian females	ethnicity x gender	FairFace_AF
Black females	ethnicity x gender	FairFace_BF
Latino females	ethnicity x gender	FairFace_LF
White females	ethnicity x gender	FairFace_WF
Asian males	ethnicity x gender	FairFace_AM
Black males	ethnicity x gender	FairFace_BM
Latino males	ethnicity x gender	FairFace_LM
White males	ethnicity x gender	FairFace_WM

**Table 2. Description of the subsets of the Chicago Face Dataset.** CFD has been split into subsets including a single gender, a single ethnicity or a single gender of a single ethnicity (column “Genders & ethnic groups”). Each subset of CFD is analyzed with each reference dataset indicated in the column “Reference datasets compared”.

<b>Genders &amp; ethnic groups</b>	<b>Test dataset identifier</b>	<b>Number of images</b>	<b>Reference datasets compared</b>
Asian females	CFD_AF	57	FairFace_All/FairFace_A/FairFace-F/ FairFace_AF
Black females	CFD_BF	104	FairFace_All/FairFace_B/FairFace-F/ FairFace_BF
Latino females	CFD_LF	56	FairFace_All/FairFace_L/FairFace-F/ FairFace_LF
White females	CFD_WF	90	FairFace_All/FairFace_W/FairFace-F/ FairFace_WF
Asian males	CFD_AM	52	FairFace_All/FairFace_A/FairFace-M/ FairFace_AM
Black males	CFD_BM	93	FairFace_All/FairFace_B/FairFace-M/ FairFace_BM
Latino males	CFD_LM	52	FairFace_All/FairFace_L/FairFace-M/ FairFace_LM
White males	CFD_WM	93	FairFace_All/FairFace_W/FairFace-M/

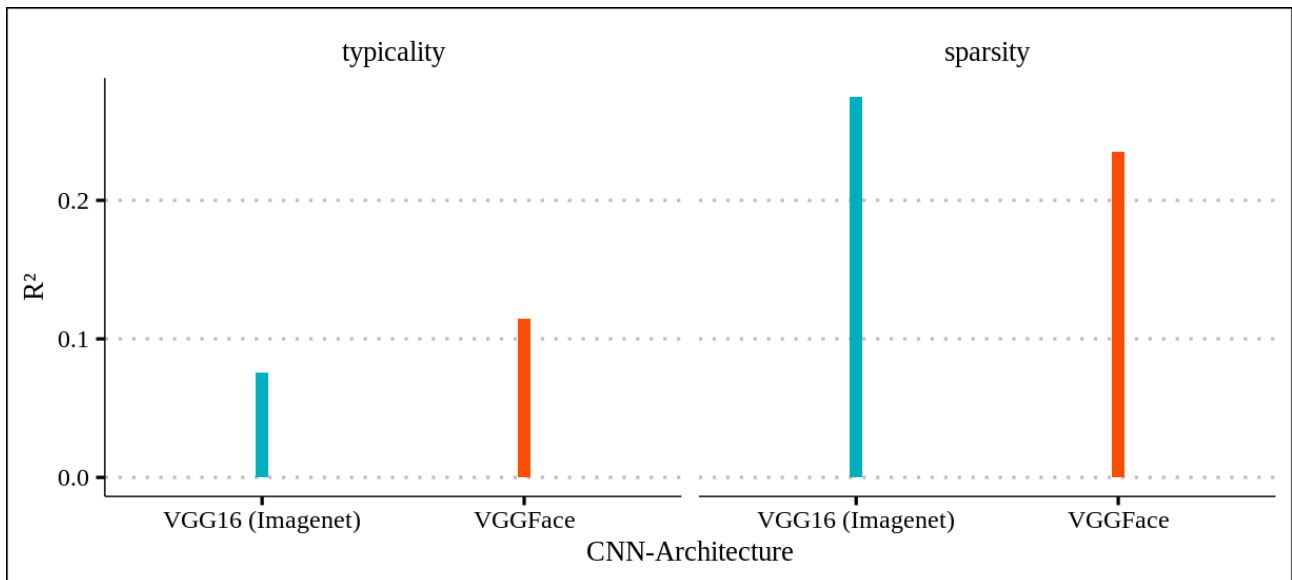
## RESULTS

### Comparing statistical typicality and sparsity

We first evaluated the ability of statistical typicality, measured from the LLH of face encodings at each layer of the CNN, to explain variation in facial attractiveness. Using portrait images of FairFace as a reference dataset to calculate the LLHs of images of the CFD dataset, we found that statistical typicality explains 8% of variance in facial attractiveness ( $R^2$ ) with VGG16, and 11% with VGGFace (Figure 3A).

In comparison, sparsity accounted for 27% and 23% of the variance in attractiveness with VGG16 and VGGFace, respectively (Figure 3B). These results suggest that variation in the sparsity of neuronal activations triggered by facial features explain variation in attractiveness more than the statistical typicality of these facial features does.

We then investigated whether we could increase the explanatory capacities of statistical typicality by including LLHs from a subset of VGGFace layers, the CNN that yielded the highest  $R^2$  for this fluency metric, rather than considering all layers. We considered four subsets of layers: regression models with 'LastConvLayer' include the last convolutional layer of each block, 'PoolingLayers' models includes all pooling layers, 'ShallowLayers', 'MiddleLayers' and 'DeepLayers' models include early, mid-tier and the deeper layers of the network, respectively. The specific layers included in each of these subsets are detailed in Figure S1. With none of these subsets the  $R^2$  score surpassed the one reached by sparsity. More precisely, we found that the fraction of explained variance was lower compared to when considering all layers, except with the 'DeepLayers' subset that yielded similar performances (11% of explained variance of attractiveness; Figure S1).



**Figure 3.** Comparison of explained variance of attractiveness ( $R^2$ ) using sparsity and statistical typicality metrics derived from VGG16(Imagenet) and VGGFace architectures.

### **Influence of reference dataset specialization**

In the previous analysis of statistical typicality, we found a slightly higher explanatory power when using VGGFace compared to VGG16 with the weights of Imagenet. This could be due to the specialization of VGGFace to process faces, leading to PDFs that are more tightly tuned to facial features and thus to more meaningful values of LLH. We thus wondered whether the statistical typicality metric would be more powerful when specializing the PDFs even further, such that LLHs are calculated in reference to one gender or one ethnic group only, or even one gender of one ethnic group, rather than all faces considered together.

To investigate the role of specializing the reference datasets and PDFs further, we performed a regression model with the  $R^2$  of the 15 models with different levels of specialization (Figure S2) as a response variable. When analyzing the different levels of the categorical variable “Reference specialization type”, we found that 'ethnicity x gender' (estimate = -0.03, 95% CI [-0.06, 0],  $p = 0.027$ ; Table 2) significantly but negatively influenced the explanatory power of statistical typicality. This result indicates that specializing the reference dataset and associated PDFs does not increase the

ability of statistical typicality to explain variation in facial attractiveness. On the contrary, we obtained the best performance when using the All reference dataset, that is, when images were sampled across all genders and ethnic groups of the FairFace dataset. Importantly, variation in  $R^2$  is not due to differences in sample size, which were kept constant across datasets (52 images).

**Table 2. Influence of reference dataset specialization on facial attractiveness prediction.** Results from GLMM analysis exploring how different reference specialization types - 'ethnicity', 'gender', and 'ethnicity x Gender' - influence the accuracy of predicting facial attractiveness. Fixed effects include the number of photos from CFD categories, while random effects encompass the 8 *Categories\_“ethnicity x gender”* from CFD images (as detailed in Table 1). The 'All' reference specialization type, with diverse images across genders and ethnicities, demonstrates significantly better predictive accuracy than 'ethnicity' and 'ethnicity x gender' reference specialization types.

<b>Predictors</b>	<b>Estimate s</b>	<b>CI</b>	<b>p</b>
(Intercept)	0.32	0.28 - 0.35	<b>&lt;0.001</b>
Reference specialization type [ethnicity] (ref: All)	-0.02	-0.05 - 0.01	0.135
Reference specialization type [Gender] (ref: All)	-0.02	-0.05 - 0.01	0.112
Reference specialization type [ethnicity x gender] (ref:All)	-0.03	-0.06 - 0.00	<b>0.027</b>
<b>Random effects</b>			
N <i>CFD_Categories_“ethnicity x gender”</i>	8		
Observations	32		
Marginal $R^2$ / Conditional $R^2$	0.044 / 0.593		

## DISCUSSION

The objective of this study was to compare two metrics of fluency - one of the main determinants of beauty - in predicting facial attractiveness. By comparing statistical models performed with the same dataset and including

the same number of explanatory variables, we showed that the sparsity of neuronal activations is a better predictor than the statistical typicality of features.

This result sheds light on the neural mechanisms involved in the sensation of fluency studied in psychology. By analyzing the different types of preferences involved in fluency, Renoult and Mendelson [13] proposed that a stimulus is fluent when it is processed either with efficacy by the brain, meaning with high-quality processing and therefore minimal information loss, or with efficiency, meaning optimal resource utilization. The question of whether it is the efficiency or rather the efficacy of processing that explains the attractiveness of such stimuli has remained unaddressed. For example, it has been shown that prototypes are attractive because they are easy for the mind [37], but ease can equally describe efficiency and efficacy. Prototypes are effective stimuli because they are most quickly and precisely categorized and stored the longest in memory [37]. However, prototypes are also efficient because they only need to stimulate a few highly selective neurons to be recognized [38]. Sparsity measured through the Gini index characterizes efficiency. However, like typicality statistical typicality measured through LLH can describe either efficacy, efficiency, or both. Accordingly, we found that across CNN's layers statistically typical faces are also sparsely encoded (mean Pearson correlation between sparsity and LLH:  $R=0,4$  for VGG,  $R=0,15$  for VGGFace). Our results thus provide evidence that fluency corresponds to processing efficiency more than processing efficacy.

The correlation between LLH and sparsity is low to moderate, though, and thus it is possible some information about attractiveness conveyed by statistical typicality is not accounted by neuronal sparsity. Dibot et al. [23] analyzed the contribution of the sparsity of individual CNN's layers and found that the first layers of VGG16-Imagenet explained most of the variance in facial beauty. This result is consistent with that of Renoult et al. [19], who showed that activation sparsity in a model of the primary visual cortex explains up to 17% of the variance. Indeed in VGG16 the first convolutional layers have been shown to model feature extraction as it operates in the primary visual cortex [17]. In contrast, with statistical typicality we found that



the fraction of explained variance was the highest for the deeper layers (11%), and adding information from the shallower layers did not increase the  $R^2$  score further. In a CNN, the deeper layers encode complex patterns and their arrangement at the largest spatial level; for faces, this means the shape and coloration of facial elements (i.e. the nose, the mouth, the eyes) but also their relative position in the face [39]. At this configural level of perception, high statistical typicality thus likely increases the ability to recognize a face individually and to memorize it. Thus, while fluency could be driven by the efficiency of information processing in the early stages of the visual system, it may be more strongly influenced by efficacy in the later stages of the processing pathway. It is noteworthy that, if they affect different stages of the information processing pathway, efficacy and efficiency could nevertheless both influence the overall fluency. Indeed, previous studies in psychology have shown that the ease of processing information at different stage of the visual system, as measured by detection time in the early stages and recognition performance in the later stages, triggers micro-experiences of fluency that aggregate into one global sensation of fluency [5,40]. To investigate further whether and how sparsity and typicality differently influence fluency at different stages of visual perception, future studies should aim to correlate these metrics with empirical measures of fluency in detection and recognition tasks.

One previous study analyzed facial attractiveness using the CFD dataset and a metric of statistical typicality based on Log-Likelihood (LLH) [24]. However, while we employed CNNs to model perception, this study used an alternative method (AAM). The authors found a strong correlation ( $r=0.386$ , corresponding to  $R^2 = 0,15$ ) between attractiveness ratings and the LLH of the stimuli. Their model integrates both low (texture) and high-level (global form) information, analogous to the deeper layers of our network. These results roughly align with our own results ( $R^2 = 0,15$  mostly explained by the deepest layers), confirming the importance of configural perception and statistical typicality when evaluating facial beauty.

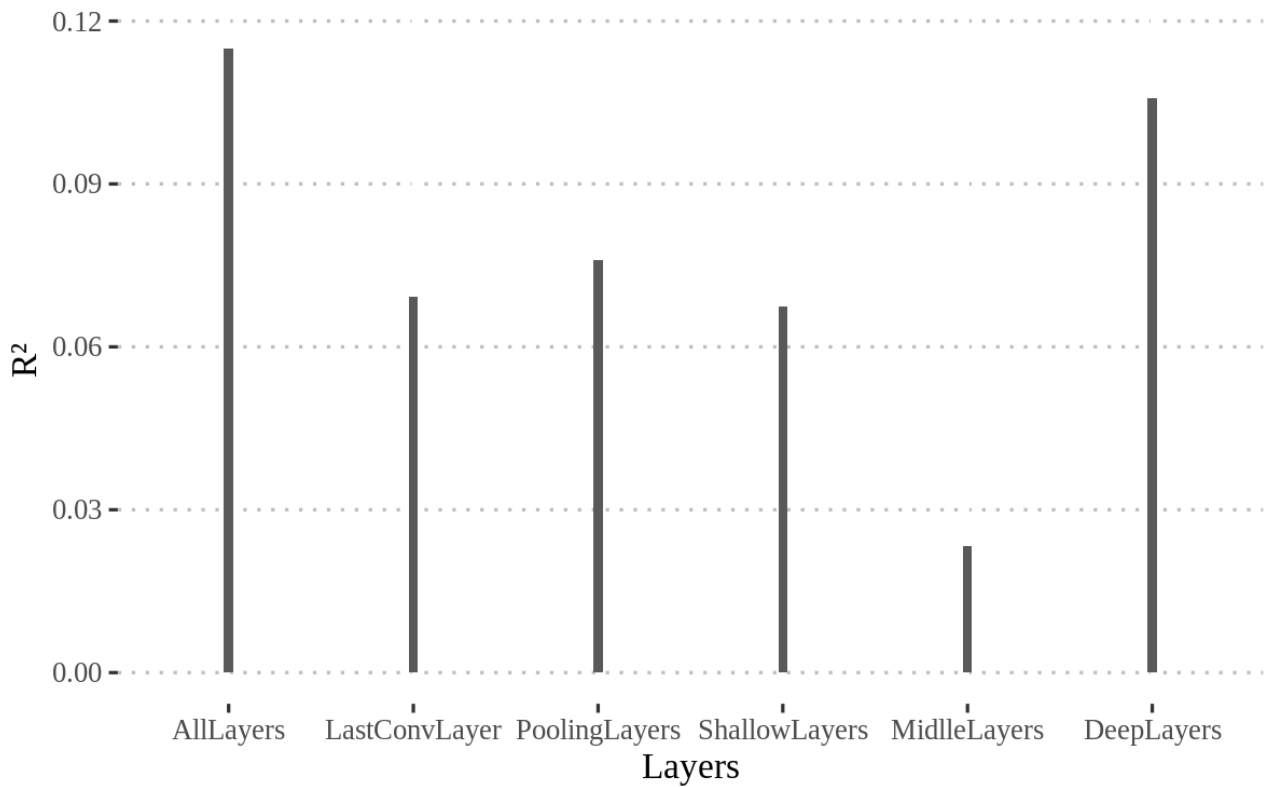
To give statistical typicality the maximum of chances, we explored the possibility of using a reference distribution refined to a specific type of face.

We incorporated these refinements based on previous results showing that the brain could interpret face signals in ethnic or gender-related subgroups [24,41–43]. We then explored different reference distributions, focused on one ethnic group, one gender, or one gender of a given ethnic group. However, we did not find evidence that refining the reference distribution increases model performances. At first glance, this result might appear contradictory to prior research indicating that facial attractiveness, when multiple groups are considered, correlates with how closely a face matches the prototype of its group. For instance, studies have found that computer-generated Caucasian and African-American faces gain attractiveness as they more closely resemble the average features of their racial group [35]. Our results may not necessarily be contradictory, though: our perceptions of beauty are deeply woven with exposure to various faces and the prototypes we form over time. Rhodes et al. (2003) emphasized that our engagements with a spectrum of faces shape our conception of the 'ideal' face [12]. Extending this notion, researchers like Lewis (2010) and Rhodes et al. (2005) posit that we might cultivate a 'composite' facial image that amalgamates average features from diverse groups, suggesting that our evaluations of attractiveness are not strictly conditioned by factors like ethnicity or gender. This hypothesis could explain why focusing the reference distribution on one specific category—be it gender, ethnicity, or their intersection—didn't enhance attractiveness predictions.

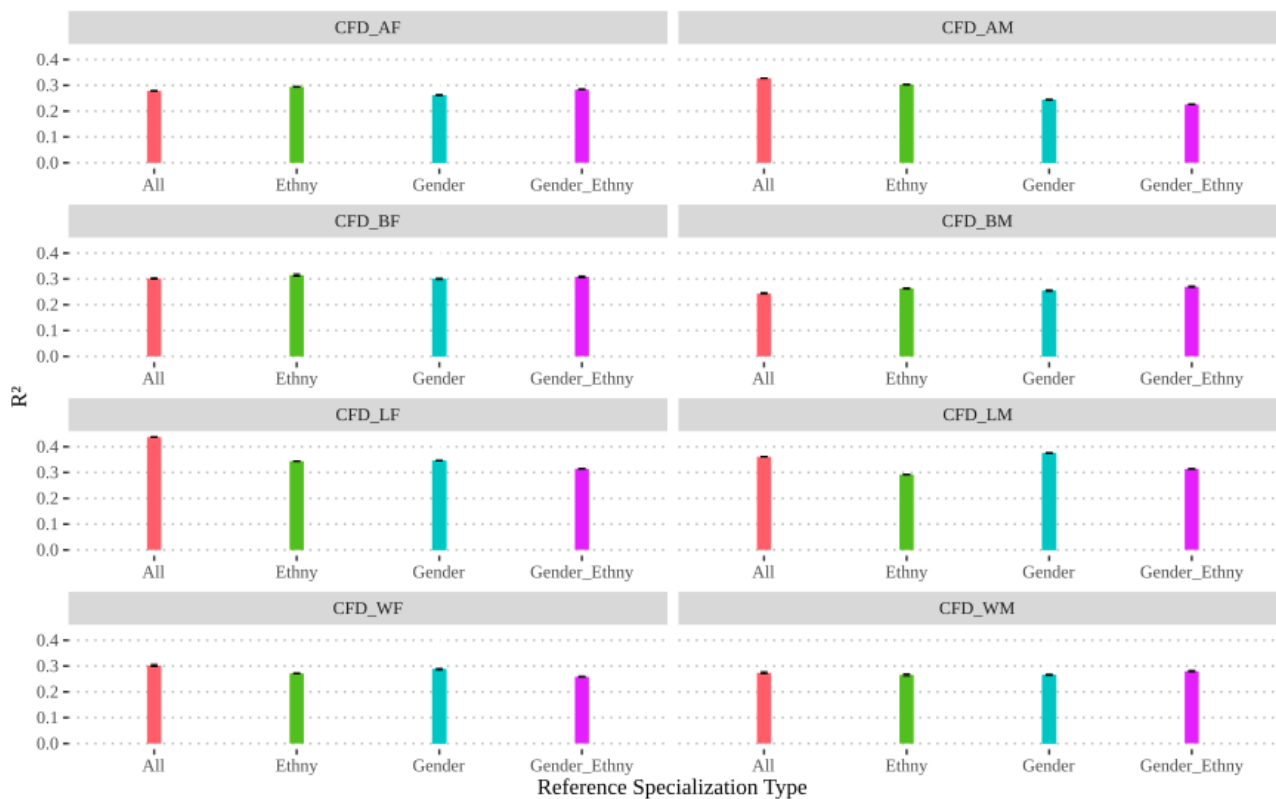
In conclusion, modeling fluency in perception involves capturing how a stimulus is processed within an observer's brain. In this research, we leveraged the potential of Convolutional Neural Networks (CNNs) to model this processing within the visual system. Although previous studies have employed CNNs to forecast beauty, they largely investigated the ability of features to directly predict attractiveness (e.g., [32]). In contrast, fluency characterizes how the visual system processes these features, thereby making it independent of the inherent meaning or nature of these features. Our use of CNNs allowed us to probe the underlying neural mechanisms of fluency. We discovered that the sparsity of neuronal activation, which portrays the efficiency of neural information processing, appears to be a more powerful determinant of beauty than statistical typicality. This finding underscores the

importance of taking into consideration the processing cost when studying attractiveness.

## **SUPPLEMENTARY DATA**



**Figure S1. Comparison of explained variance of attractiveness ( $R^2$ ) using subset of layers of VGGFace. **AllLayers:** all available layers of VGGFace. **LastConvLayer:** layers block1\_conv2, block2\_conv2, block3\_conv3, block4\_conv3, and block5\_conv3. **PoolingLayers:** layers block1\_pool, block2\_pool, block3\_pool, block4\_pool, and block5\_pool. **ShallowLayers:** layers input\_1, block1\_conv1, block1\_conv2, block1\_pool, block2\_conv1, block2\_conv2, and block2\_pool. **MiddleLayers:** Layers block3\_conv1, block3\_conv2, block3\_conv3, block3\_pool, block4\_conv1, block4\_conv2, block4\_conv3, and block4\_pool. **DeepLayers:** layers block5\_conv1, block5\_conv2, block5\_conv3, block5\_pool, fc6, fc7, fc8, and flatten.**



**Figure S2. Explained variance of attractiveness ( $R^2$ ) for the 8 Categories “ethnicity x gender” from CFD images (CFD\_AF, CFD\_AM, CFD\_BF, CFD\_BM, CFD\_LF, CFD\_LM, CFD\_WF and CFD\_WM) by reference specialization type - ‘All’ (red), ‘ethnicity’(green) , ‘gender’ (blue), and ‘ethnicity x Gender’ (pink).**

## REFERENCES

1. Batres C, Shiramizu VKM. PSA001 Secondary Analysis: Examining the “attractiveness halo effect” across cultures. PsyArXiv. 2020. doi:10.31234/osf.io/c7hf3
2. Rhodes G. The evolutionary psychology of facial beauty. *Annu Rev Psychol.* 2006;57: 199-226.
3. Rhodes G, Simmons LW, Peters M. Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evol Hum Behav.* 2005;26: 186-201.
4. Lee AY, Labroo AA. The Effect of Conceptual and Perceptual Fluency on Brand Evaluation. *J Mark Res.* 2004 [cited 12 Sep 2023]. doi:10.1509/jmkr.41.2.151.28665
5. Reber R, Schwarz N, Winkielman P. Processing fluency and aesthetic pleasure: is beauty in the perceiver’s processing experience? *Pers Soc Psychol Rev.* 2004;8: 364-382.
6. Jacobsen T, Schubotz RI, Höfel L, Cramon DY. Brain correlates of aesthetic judgment of beauty. *Neuroimage.* 2006;29.

doi:10.1016/j.neuroimage.2005.07.010

7. Reber R, Winkielman P, Schwarz N. Effects of perceptual fluency on affective judgments. *Psychol Sci.* 1998;9: 45–48.
8. Forsythe A, Nadal M, Sheehy N, Cela-Conde CJ, Sawey M. Predicting beauty: fractal dimension and visual complexity in art. *Br J Psychol.* 2011;102: 49–70.
9. Street N, Forsythe AM, Reilly R, Taylor R, Helmy MS. A Complex Story: Universal Preference vs. Individual Differences Shaping Aesthetic Response to Fractals Patterns. *Front Hum Neurosci.* 2016;10. doi:10.3389/fnhum.2016.00213
10. Winkielman P, Halberstadt J, Fazendeiro T, Catty S. Prototypes are attractive because they are easy on the mind. *Psychol Sci.* 2006;17: 799–806.
11. Winkielman P, Schwarz N, Fazendeiro TA, Reber R. The hedonic marking of processing fluency: Implications for evaluative judgment. 2003 [cited 12 Sep 2023]. Available: <https://www.semanticscholar.org/paper/The-hedonic-marking-of-processing-fluency%3A-for-Winkielman-Schwarz/750bf4a9044a127106a89bad9f90c01741f6adad>
12. Halberstadt J, Rhodes G. It's not just average faces that are attractive: computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychon Bull Rev.* 2003;10: 149–156.
13. Renoult JP, Mendelson TC. Processing bias: extending sensory drive to include efficacy and efficiency in information processing. *Proc Biol Sci.* 2019;286: 20190165.
14. Redies C. A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spat Vis.* 2007;21: 97–117.
15. Mayer S, Landwehr JR. Quantifying visual aesthetics based on processing fluency theory: Four algorithmic measures for antecedents of aesthetic preferences. *Psychology of Aesthetics, Creativity, and the Arts.* 2018 [cited 12 Sep 2023]. doi:10.1037/ACA0000187
16. Kriegeskorte N. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu Rev Vis Sci.* 2015;1: 417–446.
17. Lindsay GW. Convolutional neural networks as a model of the visual system: Past, present, and future. *J Cogn Neurosci.* 2021;33: 2017–2031.
18. Barlow HB. Possible principles underlying the transformations of sensory messages. *Sensory Communication.* The MIT Press; 1961. pp. 216–234.
19. Renoult JP, Bovet J, Raymond M. Beauty is in the efficient coding of the beholder. *R Soc Open Sci.* 2016;3: 160027.
20. Olshausen BA, Field DJ. Sparse coding of sensory inputs. *Curr Opin Neurobiol.* 2004;14: 481–487.
21. Winkielman P, Huber DE, Kavanagh L, Schwarz N. Fluency of consistency: When thoughts fit nicely and flow smoothly. *Cognitive Consistency: A Fundamental Principle in Social Cognition.* 2012; 89–111.
22. Holzleitner IJ, Lee AJ, Hahn AC, Kandrik M, Bovet J, Renoult JP, et al. Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *J Exp Psychol Hum Percept Perform.* 2019;45: 1589–1595.

23. Dibot N, Tio S Mendelson T, Puech W, Renoult J. Sparsity in an artificial neural network predicts beauty: towards a model of processing-based aesthetics. In prep.
24. Ryali CK, Goffin S, Winkielman P, Yu AJ. From likely to likable: The role of statistical typicality in human social assessment of faces. *Proc Natl Acad Sci U S A*. 2020;117: 29371–29380.
25. Briellmann AA, Dayan P. A computational model of aesthetic value. *Psychol Rev*. 2022;129: 1319–1337.
26. Ma DS, Correll J, Wittenbrink B. The Chicago face database: A free stimulus set of faces and norming data. *Behav Res Methods*. 2015;47: 1122–1135.
27. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. Available: <http://arxiv.org/abs/1409.1556>
28. Parkhi OM, Vedaldi A, Zisserman A. Deep Face Recognition. 2015 [cited 12 Sep 2023]. doi:10.5244/C.29.41
29. Güçlü U, van Gerven MAJ. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci*. 2015;35: 10005–10014.
30. Peterson JC, Abbott JT, Griffiths TL. Evaluating (and Improving) the Correspondence Between Deep Neural Networks and Human Representations. *Cogn Sci*. 2018;42: 2648–2669.
31. Hurley N, Rickard S. Comparing measures of sparsity. 2008 IEEE Workshop on Machine Learning for Signal Processing. IEEE; 2008. doi:10.1109/mlsp.2008.4685455
32. Iigaya K, Yi S, Wahle IA, Tanwisuth S, Cross L, O’Doherty JP. Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nat Commun*. 2023;14: 127.
33. Karkkainen K, Joo J. FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE; 2021. doi:10.1109/wacv48630.2021.00159
34. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 2000;42: 80.
35. Potter T, Corneille O. Locating attractiveness in the face space: faces are more attractive when closer to their group prototype. *Psychon Bull Rev*. 2008;15: 615–622.
36. Ryali CK, Yu AJ. Beauty-in-averageness and its contextual modulations: A Bayesian statistical account. *bioRxiv*. bioRxiv; 2018. doi:10.1101/360651
37. Winkielman P, Halberstadt J, Fazendeiro T, Catty S. Prototypes are attractive because they are easy on the mind. *Psychological Sciences*. 2006;17: 799–806.
38. Quiroga RQ, Kreiman G, Koch C, Fried I. Sparse but not ‘grandmother-cell’ coding in the medial temporal lobe. *Trends in Cognitive Sciences*. 2008;12: 87–91.
39. Khan K, Attique M, Khan RU, Syed I, Chung T-S. A multi-task framework for facial attributes classification through end-to-end face parsing and Deep

Convolutional Neural Networks. *Sensors* (Basel). 2020;20: 328.

40. Wurtz P, Reber R, Zimmermann TD. The feeling of fluent perception: a single experience from multiple asynchronous sources. *Conscious Cogn*. 2008;17: 171-184.
41. Levin DT. Classifying faces by race: The structure of face categories. *J Exp Psychol Learn Mem Cogn*. 1996;22: 1364-1382.
42. Kondo A, Takahashi K, Watanabe K. Influence of gender membership on sequential decisions of face attractiveness. *Atten Percept Psychophys*. 2013;75: 1347-1352.
43. Kramer RSS, Jones AL, Sharma D. Sequential effects in judgements of attractiveness: the influences of face race and sex. *PLoS One*. 2013;8: e82226.





# Bibliographie

references in citation order.

- [1] Pat MORRIS. « Animal Eyes (Oxford Animal Biology Series) – By Michael F. Land & Dan-Eric Nilsson ». In : *Zoological Journal of the Linnean Society* 166 (déc. 2012). DOI : [10.1111/j.1096-3642.2012.00849.x](https://doi.org/10.1111/j.1096-3642.2012.00849.x) (cf. p. 5).
- [2] Nancy KANWISHER et Galit YOVEL. « The fusiform face area : a cortical region specialized for the perception of faces ». In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 361.1476 (déc. 2006), p. 2109-2128. DOI : [10.1098/rstb.2006.1934](https://doi.org/10.1098/rstb.2006.1934). (Visité le 17/07/2024) (cf. p. 6).
- [3] Hanghang TONG et al. *Classification of Digital Photos Taken by Photographers or Home Users*. T. 3331. Journal Abbreviation : Lect. Notes Comput. Sci. Pages : 205 Publication Title : Lect. Notes Comput. Sci. Oct. 2004 (cf. p. 6).
- [4] Ritendra DATTA et al. « Studying Aesthetics in Photographic Images Using a Computational Approach ». en. In : *Computer Vision – ECCV 2006*. Sous la dir. d'Aleš LEONARDIS, Horst BISCHOF et Axel PINZ. Berlin, Heidelberg : Springer, 2006, p. 288-301. DOI : [10.1007/11744078\\_23](https://doi.org/10.1007/11744078_23) (cf. p. 6).
- [5] Karl GRAMMER et Randy THORNHILL. « Human (Homo sapiens) facial attractiveness and sexual selection : The role of symmetry and averageness ». In : *Journal of Comparative Psychology* 108.3 (1994). Place : US Publisher : American Psychological Association, p. 233-242. DOI : [10.1037/0735-7036.108.3.233](https://doi.org/10.1037/0735-7036.108.3.233) (cf. p. 6).
- [6] Seyed Ali AMIRSHAHI et al. « Evaluating the Rule of Thirds in Photographs and Paintings ». In : *Art & Perception* (jan. 2014). DOI : [10.1163/22134913-00002024](https://doi.org/10.1163/22134913-00002024) (cf. p. 6).
- [7] Anselm BRACHMANN et Christoph REDIES. « Computational and Experimental Approaches to Visual Aesthetics ». English. In : *Frontiers in Computational Neuroscience* 11 (nov. 2017). Publisher : Frontiers. DOI : [10.3389/fncom.2017.00102](https://doi.org/10.3389/fncom.2017.00102). (Visité le 17/07/2024) (cf. p. 7).
- [8] Stefano BALIETTI. « The human quest for discovering mathematical beauty in the arts ». In : *Proceedings of the National Academy of Sciences* 117.44 (nov. 2020). Publisher : Proceedings of the National Academy of Sciences, p. 27073-27075. DOI : [10.1073/pnas.2018652117](https://doi.org/10.1073/pnas.2018652117). (Visité le 17/07/2024) (cf. p. 7).
- [9] A. P. MØLLER. « Developmental stability and fitness : a review ». eng. In : *The American Naturalist* 149.5 (mai 1997), p. 916-932. DOI : [10.1086/286030](https://doi.org/10.1086/286030) (cf. p. 7).
- [10] Randy THORNHILL et Steven W. GANGESTAD. « Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women ». In : *Evolution and Human Behavior* 27.2 (2006). Place : Netherlands Publisher : Elsevier Science, p. 131-144. DOI : [10.1016/j.evohumbehav.2005.06.001](https://doi.org/10.1016/j.evohumbehav.2005.06.001) (cf. p. 7).
- [11] J. T MANNING, D SCUTT et D. I LEWIS-JONES. « Developmental Stability, Ejaculate Size, and Sperm Quality in Men ». In : *Evolution and Human Behavior* 19.5 (sept. 1998), p. 273-282. DOI : [10.1016/S1090-5138\(98\)00024-5](https://doi.org/10.1016/S1090-5138(98)00024-5). (Visité le 17/07/2024) (cf. p. 7).
- [12] J. T. MANNING et al. « Breast asymmetry and phenotypic quality in women ». In : *Evolution and Human Behavior* 18.4 (juill. 1997), p. 223-236. DOI : [10.1016/S0162-3095\(97\)00002-0](https://doi.org/10.1016/S0162-3095(97)00002-0). (Visité le 17/07/2024) (cf. p. 7).
- [13] R. THORNHILL et S. W. GANGESTAD. « Human facial beauty : Averageness, symmetry, and parasite resistance ». eng. In : *Human Nature (Hawthorne, N.Y.)* 4.3 (sept. 1993), p. 237-269. DOI : [10.1007/BF02692201](https://doi.org/10.1007/BF02692201) (cf. p. 7).
- [14] Coren L. APICELLA, Anthony C. LITTLE et Frank W. MARLOWE. « Facial averageness and attractiveness in an isolated population of hunter-gatherers ». eng. In : *Perception* 36.12 (2007), p. 1813-1820. DOI : [10.1068/p5601](https://doi.org/10.1068/p5601) (cf. p. 7).

- [15] Manfred MILINSKI et Theo C. M. BAKKER. « Female sticklebacks use male coloration in mate choice and hence avoid parasitized males ». en. In : *Nature* 344.6264 (mars 1990). Publisher : Nature Publishing Group, p. 330-333. DOI : [10.1038/344330a0](https://doi.org/10.1038/344330a0). (Visité le 17/07/2024) (cf. p. 7).
- [16] Sarah PRYKE et Simon GRIFFITH. « Red dominates black : agonistic signalling among head morphs in the colour polymorphic Gouldian finch ». In : *Proceedings. Biological sciences / The Royal Society* 273 (déc. 2005), p. 949-57. DOI : [10.1098/rspb.2005.3362](https://doi.org/10.1098/rspb.2005.3362) (cf. p. 7).
- [17] Joanna M. SETCHELL et E. JEAN WICKINGS. « Dominance, Status Signals and Coloration in Male Mandrills (*Mandrillus sphinx*) ». en. In : *Ethology* 111.1 (2005). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.2004.01054.x>, p. 25-50. DOI : [10.1111/j.1439-0310.2004.01054.x](https://doi.org/10.1111/j.1439-0310.2004.01054.x). (Visité le 17/07/2024) (cf. p. 7, 35, 37).
- [18] E. OTTA, F. FOLLADORE ABROSIO et R. L. HOSHINO. « Reading a smiling face : messages conveyed by various forms of smiling ». eng. In : *Perceptual and Motor Skills* 82.3 Pt 2 (juin 1996), p. 1111-1121. DOI : [10.2466/pms.1996.82.3c.1111](https://doi.org/10.2466/pms.1996.82.3c.1111) (cf. p. 7).
- [19] Ran HASSIN et Yaacov TROPE. « Facing faces : Studies on the cognitive aspects of physiognomy ». In : *Journal of Personality and Social Psychology* 78.5 (2000). Place : US Publisher : American Psychological Association, p. 837-852. DOI : [10.1037/0022-3514.78.5.837](https://doi.org/10.1037/0022-3514.78.5.837) (cf. p. 7).
- [20] Nick NEAVE et Kerry SHIELDS. « The effects of facial hair manipulation on female perceptions of attractiveness, masculinity, and dominance in male faces ». In : *Personality and Individual Differences* 45.5 (oct. 2008), p. 373-377. DOI : [10.1016/j.paid.2008.05.007](https://doi.org/10.1016/j.paid.2008.05.007). (Visité le 17/07/2024) (cf. p. 7).
- [21] Don R. OSBORN. « Beauty is as Beauty Does ? : Makeup and Posture Effects on Physical Attractiveness Judgments ». en. In : *Journal of Applied Social Psychology* 26.1 (1996). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.1996.tb01837.x>, p. 31-51. DOI : [10.1111/j.1559-1816.1996.tb01837.x](https://doi.org/10.1111/j.1559-1816.1996.tb01837.x). (Visité le 17/07/2024) (cf. p. 7).
- [22] Lisa M. DEBRUINE. « Facial resemblance increases the attractiveness of same-sex faces more than other-sex faces ». eng. In : *Proceedings. Biological Sciences* 271.1552 (oct. 2004), p. 2085-2090. DOI : [10.1098/rspb.2004.2824](https://doi.org/10.1098/rspb.2004.2824) (cf. p. 7).
- [23] Lisa M DEBRUINE. « Trustworthy but not lust-worthy : context-specific effects of facial resemblance ». In : *Proceedings of the Royal Society B : Biological Sciences* 272.1566 (mai 2005), p. 919-922. DOI : [10.1098/rspb.2004.3003](https://doi.org/10.1098/rspb.2004.3003). (Visité le 17/07/2024) (cf. p. 7).
- [24] Arthur C. DANTO. *The Transfiguration of the Commonplace*. en. URL : <https://www.hup.harvard.edu/books/9780674903463> (visité le 17/07/2024) (cf. p. 7).
- [25] Helmut LEDER et al. « A model of aesthetic appreciation and aesthetic judgments ». eng. In : *British Journal of Psychology (London, England : 1953)* 95.Pt 4 (nov. 2004), p. 489-508. DOI : [10.1348/0007126042369811](https://doi.org/10.1348/0007126042369811) (cf. p. 7).
- [26] Paul SILVIA. « Emotional Responses to Art : From Collation and Arousal to Cognition and Emotion ». In : *Review of General Psychology* 9 (déc. 2005), p. 342-357. DOI : [10.1037/1089-2680.9.4.342](https://doi.org/10.1037/1089-2680.9.4.342) (cf. p. 7).
- [27] A C LITTLE et al. « Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. » In : *Proceedings. Biological sciences / The Royal Society* 268.1462 (jan. 2001), p. 39-44. DOI : [10.1098/rspb.2000.1327](https://doi.org/10.1098/rspb.2000.1327). (Visité le 17/07/2024) (cf. p. 7).
- [28] Victor S. JOHNSTON et al. « Male facial attractiveness : evidence for hormone-mediated adaptive design ». English. In : *Evolution and Human Behavior* 4.22 (2001), p. 251-267. (Visité le 17/07/2024) (cf. p. 7, 8).
- [29] B. C. JONES et al. « Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high ». eng. In : *Hormones and Behavior* 48.3 (sept. 2005), p. 283-290. DOI : [10.1016/j.yhbeh.2005.03.010](https://doi.org/10.1016/j.yhbeh.2005.03.010) (cf. p. 7, 8).
- [30] I. S. PENTON-VOAK et al. « Menstrual cycle alters face preference ». eng. In : *Nature* 399.6738 (juin 1999), p. 741-742. DOI : [10.1038/21557](https://doi.org/10.1038/21557) (cf. p. 7, 8).

- [31] Lisa L. M. WELLING et al. « Men report stronger attraction to femininity in women's faces when their testosterone levels are high ». eng. In : *Hormones and Behavior* 54.5 (nov. 2008), p. 703-708. DOI : [10.1016/j.yhbeh.2008.07.012](https://doi.org/10.1016/j.yhbeh.2008.07.012) (cf. p. 7, 8).
- [32] A C LITTLE et al. « Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. » In : *Proceedings of the Royal Society B : Biological Sciences* 269.1496 (juin 2002), p. 1095-1100. DOI : [10.1098/rspb.2002.1984](https://doi.org/10.1098/rspb.2002.1984). (Visité le 17/07/2024) (cf. p. 7, 8).
- [33] Gillian RHODES et al. « Does sexual dimorphism in human faces signal health? » In : *Proceedings of the Royal Society of London. Series B : Biological Sciences* 270.suppl\_1 (août 2003). Publisher : Royal Society, S93-S95. DOI : [10.1098/rsbl.2003.0023](https://doi.org/10.1098/rsbl.2003.0023). (Visité le 17/07/2024) (cf. p. 7, 8).
- [34] W. D. HAMILTON. « The genetical evolution of social behaviour. I ». In : *Journal of Theoretical Biology* 7.1 (juill. 1964), p. 1-16. DOI : [10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4). (Visité le 17/07/2024) (cf. p. 7, 8).
- [35] Lisa M. DEBRUINE et al. « The health of a nation predicts their mate preferences : cross-cultural variation in women's preferences for masculinized male faces ». eng. In : *Proceedings. Biological Sciences* 277.1692 (août 2010), p. 2405-2410. DOI : [10.1098/rspb.2009.2184](https://doi.org/10.1098/rspb.2009.2184) (cf. p. 8).
- [36] Robert BROOKS et al. « National income inequality predicts women's preferences for masculinized faces better than health does ». In : *Proceedings of the Royal Society B : Biological Sciences* 278.1707 (mars 2011), p. 810-812. DOI : [10.1098/rspb.2010.0964](https://doi.org/10.1098/rspb.2010.0964). (Visité le 17/07/2024) (cf. p. 8).
- [37] Robert BORNSTEIN. « Exposure and Affect : Overview and Meta-Analysis of Research, 1968–1987 ». In : *Psychological Bulletin* 106 (sept. 1989), p. 265-289. DOI : [10.1037/0033-2909.106.2.265](https://doi.org/10.1037/0033-2909.106.2.265) (cf. p. 8).
- [38] R.B. ZAJONC. « Mere Exposure : A Gateway to the Subliminal ». en. In : *Current Directions in Psychological Science* 10.6 (déc. 2001). Publisher : SAGE Publications Inc, p. 224-228. DOI : [10.1111/1467-8721.00154](https://doi.org/10.1111/1467-8721.00154). (Visité le 17/07/2024) (cf. p. 8).
- [39] Robert B. ZAJONC et D. W. RAJECKI. « Exposure and affect : A field experiment ». en. In : *Psychonomic Science* 17.4 (oct. 1969), p. 216-217. DOI : [10.3758/BF03329178](https://doi.org/10.3758/BF03329178). (Visité le 17/07/2024) (cf. p. 8).
- [40] Anthony C. LITTLE et al. « Social learning and human mate preferences : a potential mechanism for generating and maintaining between-population diversity in attraction ». In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 366.1563 (fév. 2011), p. 366-375. DOI : [10.1098/rstb.2010.0192](https://doi.org/10.1098/rstb.2010.0192). (Visité le 17/07/2024) (cf. p. 8).
- [41] M. Dorothee AUGUSTIN, Johan WAGEMANS et Claus-Christian CARBON. « All is beautiful? Generality vs. specificity of word usage in visual aesthetics ». eng. In : *Acta Psychologica* 139.1 (jan. 2012), p. 187-201. DOI : [10.1016/j.actpsy.2011.10.004](https://doi.org/10.1016/j.actpsy.2011.10.004) (cf. p. 8).
- [42] David BUSS et Michael BARNES. « Preferences in Human Mate Selection ». In : *Journal of Personality and Social Psychology* 50 (mars 1986), p. 559-570. DOI : [10.1037/0022-3514.50.3.559](https://doi.org/10.1037/0022-3514.50.3.559) (cf. p. 8).
- [43] G. H. ELDER. « Appearance and education in marriage mobility ». eng. In : *American Sociological Review* 34.4 (août 1969), p. 519-533 (cf. p. 8).
- [44] Thomas F. CASH et Robert N. KILCULLEN. « The Aye of the Beholder : Susceptibility to Sexism and Beautyism in the Evaluation of Managerial Applicants ». en. In : *Journal of Applied Social Psychology* 15.4 (1985). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.1985.tb00903.x>, p. 591-605. DOI : [10.1111/j.1559-1816.1985.tb00903.x](https://doi.org/10.1111/j.1559-1816.1985.tb00903.x). (Visité le 17/07/2024) (cf. p. 8).
- [45] Ronald E. RIGGIO et Stanley B. WOLL. « The Role of Nonverbal Cues and Physical Attractiveness in the Selection of Dating Partners ». en. In : *Journal of Social and Personal Relationships* 1.3 (sept. 1984). Publisher : SAGE Publications Ltd, p. 347-357. DOI : [10.1177/0265407584013007](https://doi.org/10.1177/0265407584013007). (Visité le 17/07/2024) (cf. p. 8).
- [46] Ellen BERSCHIED et al. « Physical attractiveness and dating choice : A test of the matching hypothesis ». In : *Journal of Experimental Social Psychology* 7.2 (mars 1971), p. 173-189. DOI : [10.1016/0022-1031\(71\)90065-5](https://doi.org/10.1016/0022-1031(71)90065-5). (Visité le 17/07/2024) (cf. p. 8).
- [47] Elaine WALSTER et al. « Importance of physical attractiveness in dating behavior. » en. In : *Journal of Personality and Social Psychology* 4.5 (1966), p. 508-516. DOI : [10.1037/h0021188](https://doi.org/10.1037/h0021188). (Visité le 17/07/2024) (cf. p. 8).

- [48] Harold SIGALL et Nancy OSTROVE. « Beautiful but Dangerous : Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment ». In : *Journal of Personality and Social Psychology* 31 (mars 1975), p. 410-414. DOI : [10.1037/h0076472](https://doi.org/10.1037/h0076472) (cf. p. 8).
- [49] J. H. LANGLOIS et al. « Maxims or myths of beauty ? A meta-analytic and theoretical review ». eng. In : *Psychological Bulletin* 126.3 (mai 2000), p. 390-423. DOI : [10.1037/0033-2909.126.3.390](https://doi.org/10.1037/0033-2909.126.3.390) (cf. p. 8).
- [50] Karen DION, Ellen BERSCHIED et Elaine WALSTER. « What is beautiful is good. » en. In : *Journal of Personality and Social Psychology* 24.3 (1972), p. 285-290. DOI : [10.1037/h0033731](https://doi.org/10.1037/h0033731). (Visité le 17/07/2024) (cf. p. 8).
- [51] Thomas JACOBSEN. « Bridging the Arts and Sciences : A Framework for the Psychology of Aesthetics ». In : *Leonardo* 39 (avr. 2006), p. 155-162. DOI : [10.1162/leon.2006.39.2.155](https://doi.org/10.1162/leon.2006.39.2.155) (cf. p. 8, 9).
- [52] Slobodan MARKOVIĆ. « Components of aesthetic experience : aesthetic fascination, aesthetic appraisal, and aesthetic emotion ». In : *i-Perception* 3.1 (jan. 2012), p. 1-17. DOI : [10.1068/i0450aap](https://doi.org/10.1068/i0450aap). (Visité le 17/07/2024) (cf. p. 8, 9).
- [53] Christoph REDIES. « Combining universal beauty and cultural context in a unifying model of visual aesthetic experience ». English. In : *Frontiers in Human Neuroscience* 9 (avr. 2015). Publisher : Frontiers. DOI : [10.3389/fnhum.2015.00218](https://doi.org/10.3389/fnhum.2015.00218). (Visité le 17/07/2024) (cf. p. 8, 9).
- [54] Richard TAYLOR et al. « Fractals : A Resonance between Art and Nature ». In : jan. 2005, p. 53-63. DOI : [10.1007/3-540-26443-4\\_6](https://doi.org/10.1007/3-540-26443-4_6) (cf. p. 9).
- [55] Christoph REDIES, Jens HASENSTEIN et Joachim DENZLER. « Fractal-like image statistics in visual art : similarity to natural scenes ». eng. In : *Spatial Vision* 21.1-2 (2007), p. 137-148. DOI : [10.1163/156856807782753921](https://doi.org/10.1163/156856807782753921) (cf. p. 9).
- [56] Tomohiro ISHIZU et Semir ZEKI. « Toward A Brain-Based Theory of Beauty ». en. In : *PLOS ONE* 6.7 (juill. 2011). Publisher : Public Library of Science, e21852. DOI : [10.1371/journal.pone.0021852](https://doi.org/10.1371/journal.pone.0021852). (Visité le 17/07/2024) (cf. p. 9).
- [57] Rolf REBER, Norbert SCHWARZ et Piotr WINKIELMAN. « Processing Fluency and Aesthetic Pleasure : Is Beauty in the Perceiver's Processing Experience ? » en. In : *Personality and Social Psychology Review* 8.4 (nov. 2004). Publisher : SAGE Publications Inc, p. 364-382. DOI : [10.1207/s15327957pspr0804\\_3](https://doi.org/10.1207/s15327957pspr0804_3). (Visité le 17/07/2024) (cf. p. 9).
- [58] Daniel M. OPPENHEIMER. « The secret life of fluency ». In : *Trends in Cognitive Sciences* 12.6 (juin 2008), p. 237-241. DOI : [10.1016/j.tics.2008.02.014](https://doi.org/10.1016/j.tics.2008.02.014). (Visité le 17/07/2024) (cf. p. 9).
- [59] Tobias VOGEL, Moritz INGENDAHL et Piotr WINKIELMAN. « The architecture of prototype preferences : Typicality, fluency, and valence ». eng. In : *Journal of Experimental Psychology. General* 150.1 (jan. 2021), p. 187-194. DOI : [10.1037/xge0000798](https://doi.org/10.1037/xge0000798) (cf. p. 9).
- [60] Frank CÉZILLY et Dominique ALLAINÉ. « La sélection sexuelle ». In : *Biologie évolutive*. Sous la dir. de F. THOMAS, T. LEFÈVRE et M. RAYMOND. De Boeck, 2010, p. 387-422. (Visité le 17/07/2024) (cf. p. 10).
- [61] Gil G. ROSENTHAL. *Mate Choice : The Evolution of Sexual Decision Making from Microbes to Humans*. en. Google-Books-ID : XWeYDwAAQBAJ. Princeton University Press, juill. 2017 (cf. p. 10).
- [62] Steven GANGESTAD et Glenn SCHEYD. « The Evolution of Human Physical Attractiveness ». In : *Annu. Rev. Anthropol.* 34 (sept. 2005), p. 523-548. DOI : [10.1146/annurev.anthro.33.070203.143733](https://doi.org/10.1146/annurev.anthro.33.070203.143733) (cf. p. 11).
- [63] Lingyu LIANG et al. « SCUT-FBP5500 : A Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction ». In : *2018 24th International Conference on Pattern Recognition (ICPR)*. ISSN : 1051-4651. Août 2018, p. 1598-1603. DOI : [10.1109/ICPR.2018.8546038](https://doi.org/10.1109/ICPR.2018.8546038). (Visité le 17/07/2024) (cf. p. 13, 35, 36).
- [64] Sonia TIEO et al. « The Mandrillus Face Database : A portrait image database for individual and sex recognition, and age prediction in a non-human primate ». In : *Data in Brief* 47 (avr. 2023), p. 108939. DOI : [10.1016/j.dib.2023.108939](https://doi.org/10.1016/j.dib.2023.108939). (Visité le 17/07/2024) (cf. p. 13, 38, 40, 41).
- [65] K. R. CHOWDHARY. *Fundamentals of Artificial Intelligence*. Anglais. 1st ed. 2020 édition. New Delhi : Springer, India, Private Ltd, avr. 2020 (cf. p. 14).

- [66] Aurélien GÉRON. *Hands-On Machine Learning with Scikit-Learn and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems*. English. 1st edition. Beijing ; Boston : O'Reilly Media, mai 2017 (cf. p. 15-19).
- [67] Ethem ALPAYDIN. *Introduction to Machine Learning, fourth edition*. en. Google-Books-ID : uZnSDwAA-QBAJ. MIT Press, mars 2020 (cf. p. 15, 16).
- [68] Ian J. GOODFELLOW et al. *L' apprentissage profond / Ian Goodfellow, Yoshua Bengio, Aaron Courville*. français. Quantmetry, 2018. (Visité le 17/07/2024) (cf. p. 16-19).
- [69] Pratik Prabhanjan BRAHMA, Dapeng WU et Yiyuan SHE. « Why Deep Learning Works : A Manifold Disentanglement Perspective ». In : *IEEE Transactions on Neural Networks and Learning Systems* 27.10 (oct. 2016). Conference Name : IEEE Transactions on Neural Networks and Learning Systems, p. 1997-2008. DOI : [10.1109/TNNLS.2015.2496947](https://doi.org/10.1109/TNNLS.2015.2496947). (Visité le 17/07/2024) (cf. p. 17).
- [70] John BURGOYNE et Stephen MCADAMS. *Non-linear scaling techniques for uncovering the perceptual dimensions of timbre*. Journal Abbreviation : International Computer Music Conference, ICMC 2007 Publication Title : International Computer Music Conference, ICMC 2007. Jan. 2007 (cf. p. 17).
- [71] Michael ARBIB. « Warren McCulloch's Search for the Logic of the Nervous System ». In : *Perspectives in biology and medicine* 43 (fév. 2000), p. 193-216. DOI : [10.1353/pbm.2000.0001](https://doi.org/10.1353/pbm.2000.0001) (cf. p. 18).
- [72] David E. RUMELHART, Geoffrey E. HINTON et Ronald J. WILLIAMS. « Learning representations by back-propagating errors ». en. In : *Nature* 323.6088 (oct. 1986). Publisher : Nature Publishing Group, p. 533-536. DOI : [10.1038/323533a0](https://doi.org/10.1038/323533a0). (Visité le 17/07/2024) (cf. p. 18).
- [73] Alex GRAVES et al. *Connectionist temporal classification : Labelling unsegmented sequence data with recurrent neural networks*. T. 2006. Journal Abbreviation : ICML 2006 - Proceedings of the 23rd International Conference on Machine Learning Pages : 376 Publication Title : ICML 2006 - Proceedings of the 23rd International Conference on Machine Learning. Jan. 2006 (cf. p. 18).
- [74] Grace W. LINDSAY. « Convolutional Neural Networks as a Model of the Visual System : Past, Present, and Future ». eng. In : *Journal of Cognitive Neuroscience* 33.10 (sept. 2021), p. 2017-2031. DOI : [10.1162/jocn\\_a\\_01544](https://doi.org/10.1162/jocn_a_01544) (cf. p. 19, 23).
- [75] Ashish VASWANI et al. « Attention is All you Need ». In : *Advances in Neural Information Processing Systems*. T. 30. Curran Associates, Inc., 2017. (Visité le 17/07/2024) (cf. p. 20).
- [76] Pengyuan ZHOU et al. *A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming*. arXiv :2404.16038 [cs]. Jan. 2024. DOI : [10.48550/arXiv.2404.16038](https://doi.org/10.48550/arXiv.2404.16038). URL : <http://arxiv.org/abs/2404.16038> (visité le 17/07/2024) (cf. p. 20).
- [77] Kiersten M. RUFF et Rohit V. PAPPU. « AlphaFold and Implications for Intrinsically Disordered Proteins ». In : *Journal of Molecular Biology*. From Protein Sequence to Structure at Warp Speed : How Alphafold Impacts Biology 433.20 (oct. 2021), p. 167208. DOI : [10.1016/j.jmb.2021.167208](https://doi.org/10.1016/j.jmb.2021.167208). (Visité le 17/07/2024) (cf. p. 20).
- [78] Mark A. KRAMER. « Nonlinear principal component analysis using autoassociative neural networks ». en. In : *AIChE Journal* 37.2 (1991). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690370209>, p. 233-243. DOI : [10.1002/aic.690370209](https://doi.org/10.1002/aic.690370209). (Visité le 17/07/2024) (cf. p. 21).
- [79] Diederik P. KINGMA et Max WELLING. *Auto-Encoding Variational Bayes*. arXiv :1312.6114 [cs, stat]. Déc. 2022. DOI : [10.48550/arXiv.1312.6114](https://doi.org/10.48550/arXiv.1312.6114). URL : <http://arxiv.org/abs/1312.6114> (visité le 17/07/2024) (cf. p. 21).
- [80] Ian J. GOODFELLOW et al. *Generative Adversarial Networks*. arXiv :1406.2661 [cs, stat]. Juin 2014. DOI : [10.48550/arXiv.1406.2661](https://doi.org/10.48550/arXiv.1406.2661). URL : <http://arxiv.org/abs/1406.2661> (visité le 17/07/2024) (cf. p. 21).
- [81] Martin ARJOVSKY, Soumith CHINTALA et Léon BOTTOU. *Wasserstein GAN*. arXiv :1701.07875 [cs, stat]. Déc. 2017. DOI : [10.48550/arXiv.1701.07875](https://doi.org/10.48550/arXiv.1701.07875). URL : <http://arxiv.org/abs/1701.07875> (visité le 17/07/2024) (cf. p. 21).
- [82] Yann OLLIVIER, Hervé PAJOT et Cedric VILLANI. *Optimal Transport : Theory and Applications*. en. Google-Books-ID : m9EHBAAAQBAJ. Cambridge University Press, août 2014 (cf. p. 21).

- [83] Tero KARRAS, Samuli LAINE et Timo AILA. *A Style-Based Generator Architecture for Generative Adversarial Networks*. arXiv :1812.04948 [cs, stat]. Mars 2019. DOI : [10.48550/arXiv.1812.04948](https://doi.org/10.48550/arXiv.1812.04948). URL : <http://arxiv.org/abs/1812.04948> (visité le 17/07/2024) (cf. p. 22, 33, 34, 118).
- [84] A.h. BERMANO et al. « State-of-the-Art in the Architecture, Methods and Applications of StyleGAN ». en. In : *Computer Graphics Forum* 41.2 (2022). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14503>, p. 591-611. DOI : [10.1111/cgf.14503](https://doi.org/10.1111/cgf.14503). (Visité le 07/03/2024) (cf. p. 22, 33, 34, 118).
- [85] Zongze WU, Dani LISCHINSKI et Eli SHECHTMAN. *StyleSpace Analysis : Disentangled Controls for StyleGAN Image Generation*. arXiv :2011.12799 [cs]. Déc. 2020. DOI : [10.48550/arXiv.2011.12799](https://doi.org/10.48550/arXiv.2011.12799). URL : <http://arxiv.org/abs/2011.12799> (visité le 17/07/2024) (cf. p. 22, 33).
- [86] Tero KARRAS et al. *Alias-Free Generative Adversarial Networks*. arXiv :2106.12423 [cs, stat]. Oct. 2021. DOI : [10.48550/arXiv.2106.12423](https://doi.org/10.48550/arXiv.2106.12423). URL : <http://arxiv.org/abs/2106.12423> (visité le 17/07/2024) (cf. p. 22).
- [87] Jascha SOHL-DICKSTEIN et al. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. arXiv :1503.03585 [cond-mat, q-bio, stat]. Nov. 2015. DOI : [10.48550/arXiv.1503.03585](https://doi.org/10.48550/arXiv.1503.03585). URL : <http://arxiv.org/abs/1503.03585> (visité le 17/07/2024) (cf. p. 22, 118).
- [88] Stefan MAYER et Jan LANDWEHR. « Quantifying Visual Aesthetics Based on Processing Fluency Theory : Four Algorithmic Measures for Antecedents of Aesthetic Preferences ». In : *Psychology of Aesthetics, Creativity, and the Arts* 12 (oct. 2018). DOI : [10.1037/aca0000187](https://doi.org/10.1037/aca0000187) (cf. p. 22).
- [89] Kendrick N. KAY. « Principles for models of neural information processing ». eng. In : *NeuroImage* 180.Pt A (oct. 2018), p. 101-109. DOI : [10.1016/j.neuroimage.2017.08.016](https://doi.org/10.1016/j.neuroimage.2017.08.016) (cf. p. 23).
- [90] M. RIESENHUBER et T. POGGIO. « Hierarchical models of object recognition in cortex ». eng. In : *Nature Neuroscience* 2.11 (nov. 1999), p. 1019-1025. DOI : [10.1038/14819](https://doi.org/10.1038/14819) (cf. p. 23).
- [91] Nikolaus KRIEGESKORTE. « Deep neural networks : a new framework for modelling biological vision and brain information processing ». en. In : (oct. 2015). DOI : [10.1101/029876](https://doi.org/10.1101/029876). (Visité le 17/07/2024) (cf. p. 23).
- [92] Daniel L. K. YAMINS et al. « Performance-optimized hierarchical models predict neural responses in higher visual cortex ». In : *Proceedings of the National Academy of Sciences* 111.23 (juin 2014). Publisher : Proceedings of the National Academy of Sciences, p. 8619-8624. DOI : [10.1073/pnas.1403112111](https://doi.org/10.1073/pnas.1403112111). (Visité le 17/07/2024) (cf. p. 23).
- [93] Alex KRIZHEVSKY, Ilya SUTSKEVER et Geoffrey HINTON. « ImageNet Classification with Deep Convolutional Neural Networks ». In : *Neural Information Processing Systems* 25 (jan. 2012). DOI : [10.1145/3065386](https://doi.org/10.1145/3065386) (cf. p. 23).
- [94] Karen SIMONYAN et Andrew ZISSERMAN. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv :1409.1556 [cs]. Avr. 2015. DOI : [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556). URL : <http://arxiv.org/abs/1409.1556> (visité le 17/07/2024) (cf. p. 23).
- [95] Kohitij KAR et al. « Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior ». In : *Nature Neuroscience* 22.6 (2019), p. 974-983. DOI : [10.1038/s41593-019-0392-5](https://doi.org/10.1038/s41593-019-0392-5) (cf. p. 23, 24).
- [96] Michael J. RYAN et A. Stanley RAND. « Sexual Selection and Signal Evolution : The Ghost of Biases past ». In : *Philosophical Transactions : Biological Sciences* 340.1292 (1993). Publisher : Royal Society, p. 187-195. (Visité le 17/07/2024) (cf. p. 24).
- [97] Julien P. RENOULT et Tamra C. MENDELSON. « Processing bias : extending sensory drive to include efficacy and efficiency in information processing ». In : *Proceedings of the Royal Society B : Biological Sciences* 286.1900 (avr. 2019). Publisher : Royal Society, p. 20190165. DOI : [10.1098/rspb.2019.0165](https://doi.org/10.1098/rspb.2019.0165). (Visité le 17/07/2024) (cf. p. 24).
- [98] E. P. SIMONCELLI et B. A. OLSHAUSEN. « Natural image statistics and neural representation ». eng. In : *Annual Review of Neuroscience* 24 (2001), p. 1193-1216. DOI : [10.1146/annurev.neuro.24.1.1193](https://doi.org/10.1146/annurev.neuro.24.1.1193) (cf. p. 24).

- [99] Branka SPEHAR et al. « Beauty and the beholder : the role of visual sensitivity in visual preference ». English. In : *Frontiers in Human Neuroscience* 9 (sept. 2015). Publisher : Frontiers. DOI : [10.3389/fnhum.2015.00514](https://doi.org/10.3389/fnhum.2015.00514). (Visité le 17/07/2024) (cf. p. 24).
- [100] Branka SPEHAR, Nicholas WALKER et Richard P. TAYLOR. « Taxonomy of Individual Variations in Aesthetic Responses to Fractal Patterns ». English. In : *Frontiers in Human Neuroscience* 10 (juill. 2016). Publisher : Frontiers. DOI : [10.3389/fnhum.2016.00350](https://doi.org/10.3389/fnhum.2016.00350). (Visité le 17/07/2024) (cf. p. 24).
- [101] Christoph REDIES et al. « Artists portray human faces with the Fourier statistics of complex natural scenes ». eng. In : *Network (Bristol, England)* 18.3 (sept. 2007), p. 235-248. DOI : [10.1080/09548980701574496](https://doi.org/10.1080/09548980701574496) (cf. p. 24).
- [102] Piotr WINKIELMAN et al. « Fluency of consistency : When thoughts fit nicely and flow smoothly ». In : *Cognitive Consistency : A Fundamental Principle in Social Cognition*. Journal Abbreviation : Cognitive Consistency : A Fundamental Principle in Social Cognition. Jan. 2012, p. 89-111 (cf. p. 24).
- [103] Iris J. HOLZLEITNER et al. « Comparing theory-driven and data-driven attractiveness models using images of real women's faces ». eng. In : *Journal of Experimental Psychology. Human Perception and Performance* 45.12 (déc. 2019), p. 1589-1595. DOI : [10.1037/xhp0000685](https://doi.org/10.1037/xhp0000685) (cf. p. 24).
- [104] Piotr WINKIELMAN et al. « Prototypes Are Attractive Because They Are Easy on the Mind ». en. In : *Psychological Science* 17.9 (sept. 2006). Publisher : SAGE Publications Inc, p. 799-806. DOI : [10.1111/j.1467-9280.2006.01785.x](https://doi.org/10.1111/j.1467-9280.2006.01785.x). (Visité le 17/07/2024) (cf. p. 24).
- [105] Rolf REBER, Piotr WINKIELMAN et Norbert SCHWARZ. « Effects of Perceptual Fluency on Affective Judgments ». en. In : *Psychological Science* 9.1 (jan. 1998). Publisher : SAGE Publications Inc, p. 45-48. DOI : [10.1111/1467-9280.00008](https://doi.org/10.1111/1467-9280.00008). (Visité le 17/07/2024) (cf. p. 24).
- [106] Aenne A. BRIELMANN et Peter DAYAN. « A computational model of aesthetic value ». eng. In : *Psychological Review* 129.6 (nov. 2022), p. 1319-1337. DOI : [10.1037/rev0000337](https://doi.org/10.1037/rev0000337) (cf. p. 24, 25).
- [107] Bob B. M. WONG et Ulrika CANDOLIN. « How is female mate choice affected by male competition ? » eng. In : *Biological Reviews of the Cambridge Philosophical Society* 80.4 (nov. 2005), p. 559-571. DOI : [10.1017/S1464793105006809](https://doi.org/10.1017/S1464793105006809) (cf. p. 27).
- [108] Martin STEVENS. *Sensory Ecology, Behaviour, and Evolution*. en. Google-Books-ID : KmJoAgAAQBAJ. OUP Oxford, fév. 2013 (cf. p. 28).
- [109] Tamara M. FRANK. « Visual Ecology Thomas W. Cronin, Sönke Johnsen, N. Justin Marshall, and Eric J. Warrant, editors ». In : *Integrative and Comparative Biology* 55.2 (août 2015), p. 343-345. DOI : [10.1093/icb/icv069](https://doi.org/10.1093/icb/icv069). (Visité le 17/07/2024) (cf. p. 28).
- [110] Ségolène DELAITRE et al. « Female great tits (*Parus major*) reproduce earlier when paired with a male they prefer ». en. In : *Ethology* 129.9 (2023). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/eth.13381>, p. 461-471. DOI : [10.1111/eth.13381](https://doi.org/10.1111/eth.13381). (Visité le 17/07/2024) (cf. p. 28).
- [111] Dana PFEFFERLE et al. « Monkeys Spontaneously Discriminate Their Unfamiliar Paternal Kin under Natural Conditions Using Facial Cues ». In : *Current Biology* 24.15 (août 2014), p. 1806-1810. DOI : [10.1016/j.cub.2014.06.058](https://doi.org/10.1016/j.cub.2014.06.058). (Visité le 17/07/2024) (cf. p. 29-31).
- [112] E. SPELKE. « Preferential-looking methods as tools for the study of cognition in infancy. » In : 1985. (Visité le 17/07/2024) (cf. p. 29).
- [113] Sandra WINTERS, Constance DUBUC et James P. HIGHAM. « Perspectives : The Looking Time Experimental Paradigm in Studies of Animal Visual Perception and Cognition ». en. In : *Ethology* 121.7 (2015). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/eth.12378>, p. 625-640. DOI : [10.1111/eth.12378](https://doi.org/10.1111/eth.12378). (Visité le 17/07/2024) (cf. p. 29, 30, 114, 115).
- [114] Corri WAITT et al. « Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice ». eng. In : *Proceedings. Biological Sciences* 270 Suppl 2.Suppl 2 (nov. 2003), S144-146. DOI : [10.1098/rsbl.2003.0065](https://doi.org/10.1098/rsbl.2003.0065) (cf. p. 30).
- [115] Laura S. LEWIS et al. « Bonobos and chimpanzees remember familiar conspecifics for decades ». In : *Proceedings of the National Academy of Sciences* 120.52 (déc. 2023). Publisher : Proceedings of the National Academy of Sciences, e2304903120. DOI : [10.1073/pnas.2304903120](https://doi.org/10.1073/pnas.2304903120). (Visité le 17/07/2024) (cf. p. 30, 114).



- [116] Kevin A. ROSENFELD et al. « Experimental evidence that female rhesus macaques (*Macaca mulatta*) perceive variation in male facial masculinity ». In : *Royal Society Open Science* 6.1 (jan. 2019), p. 181415. DOI : [10.1098/rsos.181415](https://doi.org/10.1098/rsos.181415). (Visité le 17/07/2024) (cf. p. 30, 31, 35).
- [117] Vojtěch FIALA et al. « Facial attractiveness and preference of sexual dimorphism : A comparison across five populations ». en. In : *Evolutionary Human Sciences* 3 (jan. 2021), e38. DOI : [10.1017/ehs.2021.33](https://doi.org/10.1017/ehs.2021.33). (Visité le 17/07/2024) (cf. p. 31).
- [118] Sandra WINTERS, William L. ALLEN et James P. HIGHAM. *The structure of species discrimination signals across a primate radiation*. en. Pages : 574558 Section : New Results. Avr. 2019. DOI : [10.1101/574558](https://doi.org/10.1101/574558). URL : <https://www.biorxiv.org/content/10.1101/574558v2> (visité le 17/07/2024) (cf. p. 31-33, 35).
- [119] Christophe A. H. BOUSQUET et Gwenaël KAMINSKI. « Transforming faces to mimic natural kin : A comparison of different paradigms ». en. In : *Behavior Research Methods* 54.1 (fév. 2022), p. 13-25. DOI : [10.3758/s13428-021-01614-5](https://doi.org/10.3758/s13428-021-01614-5). (Visité le 17/07/2024) (cf. p. 31).
- [120] Sarah NILA et al. « Male Homosexual Preference : Femininity and the Older Brother Effect in Indonesia ». eng. In : *Evolutionary Psychology : An International Journal of Evolutionary Approaches to Psychology and Behavior* 17.4 (2019), p. 1474704919880701. DOI : [10.1177/1474704919880701](https://doi.org/10.1177/1474704919880701) (cf. p. 31-33).
- [121] Duncan ROWLAND et David PERRETT. « Manipulating Facial Appearance Through Shape and Color ». In : *Computer Graphics and Applications, IEEE* 15 (oct. 1995), p. 70-76. DOI : [10.1109/38.403830](https://doi.org/10.1109/38.403830) (cf. p. 31, 32).
- [122] Elad RICHARDSON et al. *Encoding in Style : a StyleGAN Encoder for Image-to-Image Translation*. arXiv :2008.00951 [cs]. Avr. 2021. DOI : [10.48550/arXiv.2008.00951](https://doi.org/10.48550/arXiv.2008.00951). URL : <http://arxiv.org/abs/2008.00951> (visité le 17/07/2024) (cf. p. 33, 34, 118).
- [123] Sergio SOLARI et Robert J. BAKER. « Mammal Species of the World : A Taxonomic and Geographic Reference by D. E. Wilson ; D. M. Reeder ». In : *Journal of Mammalogy* 88.3 (juin 2007), p. 824-830. DOI : [10.1644/06-MAMM-R-422.1](https://doi.org/10.1644/06-MAMM-R-422.1). (Visité le 17/07/2024) (cf. p. 35).
- [124] Debbie S. MA, Justin KANTNER et Bernd WITTENBRINK. « Chicago Face Database : Multiracial expansion ». en. In : *Behavior Research Methods* 53.3 (juin 2021), p. 1289-1300. DOI : [10.3758/s13428-020-01482-5](https://doi.org/10.3758/s13428-020-01482-5). (Visité le 17/07/2024) (cf. p. 35, 36).
- [125] Joanna M. SETCHELL, Marie CHARPENTIER et E. Jean WICKINGS. « Sexual Selection and Reproductive Careers in Mandrills (*Mandrillus sphinx*) ». In : *Behavioral Ecology and Sociobiology* 58.5 (2005). Publisher : Springer, p. 474-485. (Visité le 17/07/2024) (cf. p. 35, 37, 38, 40).
- [126] Marie JE CHARPENTIER et al. « Mandrill mothers associate with infants who look like their own offspring using phenotype matching ». In : *eLife* 11 (nov. 2022). Sous la dir. d'Ashleigh S GRIFFIN, Christian RUTZ et James P HIGHAM. Publisher : eLife Sciences Publications, Ltd, e79417. DOI : [10.7554/eLife.79417](https://doi.org/10.7554/eLife.79417). (Visité le 07/03/2024) (cf. p. 35, 39, 40).
- [127] Joanna M SETCHELL, E JEAN WICKINGS et Leslie A KNAPP. « Signal content of red facial coloration in female mandrills (*Mandrillus sphinx*) ». In : *Proceedings of the Royal Society B : Biological Sciences* 273.1599 (sept. 2006), p. 2395-2400. DOI : [10.1098/rspb.2006.3573](https://doi.org/10.1098/rspb.2006.3573). (Visité le 17/07/2024) (cf. p. 35, 113).
- [128] Sonia TIEO et al. « Social and sexual consequences of facial femininity in a non-human primate ». eng. In : *iScience* 26.10 (oct. 2023), p. 107901. DOI : [10.1016/j.isci.2023.107901](https://doi.org/10.1016/j.isci.2023.107901) (cf. p. 35, 37, 39, 40, 71, 114).
- [129] K. A. ABERNETHY, L. J. T. WHITE et E. J. WICKINGS. « Hordes of mandrills (*Mandrillus sphinx*) : extreme group size and seasonal male presence ». en. In : *Journal of Zoology* 258.1 (2002). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1017/S0952836902001267>, p. 131-137. DOI : [10.1017/S0952836902001267](https://doi.org/10.1017/S0952836902001267). (Visité le 17/07/2024) (cf. p. 37-39).
- [130] Joanna M. SETCHELL. « Behavioural Development in Male Mandrills (*Mandrillus sphinx*) : Puberty to Adulthood ». In : *Behaviour* 140.8/9 (2003). Publisher : Brill, p. 1053-1089. (Visité le 17/07/2024) (cf. p. 37, 38).

- [131] J. M. SETCHELL et A. F. DIXSON. « Changes in the secondary sexual adornments of male mandrills (*Mandrillus sphinx*) are associated with gain and loss of alpha status ». eng. In : *Hormones and Behavior* 39.3 (mai 2001), p. 177-184. DOI : [10.1006/hbeh.2000.1628](https://doi.org/10.1006/hbeh.2000.1628) (cf. p. 37, 38).
- [132] Charles DARWIN, John Tyler BONNER et Robert M. MAY. *The Descent of Man, and Selection in Relation to Sex*. REV - Revised. Princeton University Press, 1981. (Visité le 17/07/2024) (cf. p. 37-39).
- [133] Joanna M. SETCHELL et E. JEAN WICKINGS. « Mate Choice in Male Mandrills (*Mandrillus sphinx*) ». en. In : *Ethology* 112.1 (2006). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.2006.01128.x>, p. 91-99. DOI : [10.1111/j.1439-0310.2006.01128.x](https://doi.org/10.1111/j.1439-0310.2006.01128.x). (Visité le 17/07/2024) (cf. p. 37, 39, 40).
- [134] Patricia PEIGNOT et al. « Learning from the first release project of captive-bred mandrills *Mandrillus sphinx* in Gabon ». en. In : *Oryx* 42.1 (jan. 2008), p. 122-131. DOI : [10.1017/S0030605308000136](https://doi.org/10.1017/S0030605308000136). (Visité le 17/07/2024) (cf. p. 37, 39).
- [135] Julien P. RENOULT et al. « The Evolution of the Multicoloured Face of Mandrills : Insights from the Perceptual Space of Colour Vision ». In : *PLoS ONE* 6.12 (déc. 2011), e29117. DOI : [10.1371/journal.pone.0029117](https://doi.org/10.1371/journal.pone.0029117). (Visité le 17/07/2024) (cf. p. 39, 40).
- [136] Nicolas M. DIBOT et al. « Sparsity in an artificial neural network predicts beauty : Towards a model of processing-based aesthetics ». en. In : *PLOS Computational Biology* 19.12 (déc. 2023). Publisher : Public Library of Science, e1011703. DOI : [10.1371/journal.pcbi.1011703](https://doi.org/10.1371/journal.pcbi.1011703). (Visité le 07/03/2024) (cf. p. 42).
- [137] Anthony C. LITTLE et al. « Men's strategic preferences for femininity in female faces ». en. In : *British Journal of Psychology* 105.3 (2014). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12043>, p. 364-381. DOI : [10.1111/bjop.12043](https://doi.org/10.1111/bjop.12043). (Visité le 17/07/2024) (cf. p. 111).
- [138] Sarah BANCHEFSKY et al. « But You Don't Look Like A Scientist ! : Women Scientists with Feminine Appearance are Deemed Less Likely to be Scientists ». In : *Sex Roles* 75 (août 2016). DOI : [10.1007/s11199-016-0586-1](https://doi.org/10.1007/s11199-016-0586-1) (cf. p. 111).
- [139] Sabine SCZESNY, Sandra SPREEMANN et Dagmar STAHLBERG. « Masculine = Competent ? Physical Appearance and Sex as Sources of Gender-Stereotypic Attributions ». In : *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie* 65 (mars 2006), p. 15-23. DOI : [10.1024/1421-0185.65.1.15](https://doi.org/10.1024/1421-0185.65.1.15) (cf. p. 111).
- [140] Robert L. FANTZ. « Visual Experience in Infants : Decreased Attention to Familiar Patterns Relative to Novel Ones ». In : *Science* 146.3644 (oct. 1964). Publisher : American Association for the Advancement of Science, p. 668-670. DOI : [10.1126/science.146.3644.668](https://doi.org/10.1126/science.146.3644.668). (Visité le 17/07/2024) (cf. p. 114).
- [141] Kazuo FUJITA. « Species recognition by five macaque monkeys ». en. In : *Primates* 28.3 (juill. 1987), p. 353-366. DOI : [10.1007/BF02381018](https://doi.org/10.1007/BF02381018). (Visité le 17/07/2024) (cf. p. 114).
- [142] Jenny RICHMOND, Michael COLOMBO et Harlene HAYNE. « Interpreting visual preferences in the visual paired-comparison task ». eng. In : *Journal of Experimental Psychology. Learning, Memory, and Cognition* 33.5 (sept. 2007), p. 823-831. DOI : [10.1037/0278-7393.33.5.823](https://doi.org/10.1037/0278-7393.33.5.823) (cf. p. 114).
- [143] Junghyun PARK, Eiko SHIMOJO et Shinsuke SHIMOJO. « Roles of familiarity and novelty in visual preference judgments are segregated across object categories ». In : *Proceedings of the National Academy of Sciences* 107.33 (août 2010). Publisher : Proceedings of the National Academy of Sciences, p. 14552-14555. DOI : [10.1073/pnas.1004374107](https://doi.org/10.1073/pnas.1004374107). (Visité le 17/07/2024) (cf. p. 114).
- [144] M. HOFSTADTER et J. S. REZNICK. « Response modality affects human infant delayed-response performance ». eng. In : *Child Development* 67.2 (avr. 1996), p. 646-658 (cf. p. 115).
- [145] A. AHMED et T. RUFFMAN. « Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task ? » eng. In : *Developmental Psychology* 34.3 (mai 1998), p. 441-453. DOI : [10.1037//0012-1649.34.3.441](https://doi.org/10.1037//0012-1649.34.3.441) (cf. p. 115).
- [146] Eric P. CHARLES et Susan M. RIVERA. « Object permanence and method of disappearance : looking measures further contradict reaching measures ». eng. In : *Developmental Science* 12.6 (nov. 2009), p. 991-1006. DOI : [10.1111/j.1467-7687.2009.00844.x](https://doi.org/10.1111/j.1467-7687.2009.00844.x) (cf. p. 115).

- [147] R. S. BOGARTZ, J. L. SHINSKEY et C. J. SPEAKER. « Interpreting infant looking : the event set x event set design ». eng. In : *Developmental Psychology* 33.3 (mai 1997), p. 408-422. DOI : [10.1037//0012-1649.33.3.408](https://doi.org/10.1037//0012-1649.33.3.408) (cf. p. 115).
- [148] Alice BANIEL, Guy COWLISHAW et Elise HUCHARD. « Male violence and sexual intimidation in a wild primate society ». In : *Current biology* 27.14 (2017). Publisher : Elsevier, p. 2163-2168. (Visité le 17/07/2024) (cf. p. 116).
- [149] Manuel BRACK et al. *LEDITS++ : Limitless Image Editing using Text-to-Image Models*. arXiv :2311.16711 [cs]. Juin 2024. DOI : [10.48550/arXiv.2311.16711](https://doi.org/10.48550/arXiv.2311.16711). URL : <http://arxiv.org/abs/2311.16711> (visité le 17/07/2024) (cf. p. 118).
- [150] Chen-Yi LU, Dan Jeric ARCEGA RUSTIA et Ta-Te LIN. « Generative Adversarial Network Based Image Augmentation for Insect Pest Classification Enhancement ». In : *IFAC-PapersOnLine*. 6th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2019 52.30 (jan. 2019), p. 1-5. DOI : [10.1016/j.ifacol.2019.12.406](https://doi.org/10.1016/j.ifacol.2019.12.406). (Visité le 17/07/2024) (cf. p. 119).
- [151] Ali JAMALI et al. « 3DUNetGSFormer : A deep learning pipeline for complex wetland mapping using generative adversarial networks and Swin transformer ». In : *Ecological Informatics* 72 (déc. 2022), p. 101904. DOI : [10.1016/j.ecoinf.2022.101904](https://doi.org/10.1016/j.ecoinf.2022.101904). (Visité le 17/07/2024) (cf. p. 119).
- [152] Simon L. MADSEN et al. « Generating artificial images of plant seedlings using generative adversarial networks ». In : *Biosystems Engineering* 187 (nov. 2019), p. 147-159. DOI : [10.1016/j.biosystemseng.2019.09.005](https://doi.org/10.1016/j.biosystemseng.2019.09.005). (Visité le 17/07/2024) (cf. p. 119).
- [153] Jia WANG et Shigeru TABETA. « Four-channel generative adversarial networks can predict the distribution of reef-associated fish in the South and East China Seas ». In : *Ecological Informatics* 78 (déc. 2023), p. 102321. DOI : [10.1016/j.ecoinf.2023.102321](https://doi.org/10.1016/j.ecoinf.2023.102321). (Visité le 17/07/2024) (cf. p. 119, 121).
- [154] Wakinyan BENHAMOU et al. « Phenotypic evolution of SARS-CoV-2 : a statistical inference approach ». eng. In : *Evolution; International Journal of Organic Evolution* 77.10 (oct. 2023), p. 2213-2223. DOI : [10.1093/evolut/qpad133](https://doi.org/10.1093/evolut/qpad133) (cf. p. 119).
- [155] Uri HASSON, Samuel A. NASTASE et Ariel GOLDSTEIN. « Direct Fit to Nature : An Evolutionary Perspective on Biological and Artificial Neural Networks ». eng. In : *Neuron* 105.3 (fév. 2020), p. 416-434. DOI : [10.1016/j.neuron.2019.12.002](https://doi.org/10.1016/j.neuron.2019.12.002) (cf. p. 119).
- [156] Yseult HÉJJA-BRICHARD et al. « Using generative artificial intelligence to test hypotheses about animal signal evolution : A case study in an ornamented fish ». In : (2023). (Visité le 17/07/2024) (cf. p. 120).
- [157] Laszlo TALAS et al. « CamoGAN : Evolving optimum camouflage with Generative Adversarial Networks ». en. In : *Methods in Ecology and Evolution* 11.2 (2020). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13334>, p. 240-247. DOI : [10.1111/2041-210X.13334](https://doi.org/10.1111/2041-210X.13334). (Visité le 17/07/2024) (cf. p. 120, 121).
- [158] Anthony ARAK et al. « Hidden preferences and the evolution of signals ». In : *Philosophical Transactions of the Royal Society of London. Series B : Biological Sciences* 340.1292 (jan. 1997). Publisher : Royal Society, p. 207-213. DOI : [10.1098/rstb.1993.0059](https://doi.org/10.1098/rstb.1993.0059). (Visité le 17/07/2024) (cf. p. 120, 121).
- [159] Irun R. COHEN et Assaf MARRON. « Evolution is driven by natural autoencoding : reframing species, interaction codes, cooperation and sexual reproduction ». In : *Proceedings of the Royal Society B : Biological Sciences* 290.1994 (mars 2023). Publisher : Royal Society, p. 20222409. DOI : [10.1098/rspb.2022.2409](https://doi.org/10.1098/rspb.2022.2409). (Visité le 17/07/2024) (cf. p. 121).
- [160] Johannes HIRN et al. « A deep Generative Artificial Intelligence system to predict species coexistence patterns ». en. In : *Methods in Ecology and Evolution* 13.5 (2022). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13827>, p. 1052-1061. DOI : [10.1111/2041-210X.13827](https://doi.org/10.1111/2041-210X.13827). (Visité le 17/07/2024) (cf. p. 121).
- [161] Amédée ROY, Ronan FABLET et Sophie Lanco BERTRAND. « Using generative adversarial networks (GAN) to simulate central-place foraging trajectories ». en. In : *Methods in Ecology and Evolution* 13.6 (2022). \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13853>, p. 1275-1287. DOI : [10.1111/2041-210X.13853](https://doi.org/10.1111/2041-210X.13853). (Visité le 17/07/2024) (cf. p. 121).





# Résumé

Les visages, cet ensemble de signaux visuels constitués de couleurs, de motifs, de textures et de mouvements, jouent un rôle fondamental dans les interactions entre les individus, chez de nombreuses espèces. La question de l'origine des préférences pour certains visages plutôt que pour d'autres est complexe et peut être explorée à l'aune de plusieurs disciplines : sciences de l'évolution, écologie, neurosciences, psychologie ou encore philosophie. Nous montrons dans cette thèse comment l'intelligence artificielle peut créer un pont entre ces différentes disciplines.

Le Projet Mandrillus, projet de terrain de long-terme situé au sud du Gabon, étudie depuis plus de 10 ans cette communication au sein d'une population de 250 mandrills (*Mandrillus sphinx*), un singe des forêts équatoriales d'Afrique Centrale. Des travaux récents ont notamment montré une corrélation entre la féminité des visages des mandrills et leur attractivité pour des congénères. Alors que chez les humains, ce lien est positif, chez les mandrills il est négatif.

Un premier objectif de cette thèse est de valider expérimentalement ces résultats, qui n'ont pour le moment été obtenus qu'avec des études corrélatives. En conditions contrôlées avec des mandrills en captivité, nous avons testé l'attractivité d'images générées artificiellement et variant selon un paramètre de féminité, qui ont confirmé et renforcé ce lien négatif entre féminité et attractivité.

Pour faire varier le paramètre de féminité, nous avons eu recours à l'intelligence artificielle (IA) générative, qui permet de créer des images artificielles à partir d'images réelles en faisant varier certaines caractéristiques spécifiques. Le développement méthodologique d'une approche d'IA générative applicable à des expériences comportementales, avec les mandrills ou d'autres espèces animales, constitue le deuxième objectif de la thèse.

Au-delà de la synthèse et la manipulation de stimuli visuels, l'IA peut éclairer les processus communicatifs entre animaux, ainsi que leur évolution, de part ses similarités structurelles avec la perception visuelle biologique. Certains réseaux de neurones artificiels convolutifs permettent ainsi de prédire l'activité du cortex visuel, et de mieux comprendre le traitement cérébral de l'information. Le troisième objectif est de montrer comment un modèle basé sur ce type d'approche peut prédire la beauté chez les humains.

