



HAL
open science

Utilisation de réseaux de neurones convolutifs pour mieux comprendre l'évolution et le développement du comportement de dessin chez les Hominidés

Benjamin Beltzung

► **To cite this version:**

Benjamin Beltzung. Utilisation de réseaux de neurones convolutifs pour mieux comprendre l'évolution et le développement du comportement de dessin chez les Hominidés. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université de Strasbourg, 2023. Français. NNT : 2023STRAJ114 . tel-04780844

HAL Id: tel-04780844

<https://theses.hal.science/tel-04780844v1>

Submitted on 13 Nov 2024

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

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



ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

CNRS IPHC UMR 7178

THÈSE

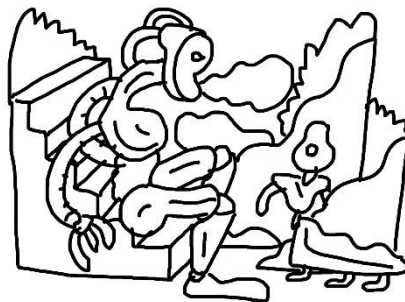
présentée par

Benjamin Beltzung

soutenue le : 17 octobre 2023

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Bioinformatique



Utilisation de réseaux de neurones convolutifs pour mieux comprendre l'évolution et le développement du comportement de dessin chez les Hominidés

THÈSE dirigée par :
SUEUR Cédric

Maître de conférences, Université de Strasbourg

RAPPORTEURS :
SOULÉ-DUPUY Chantal
TESTE Olivier

Professeure des universités, Université Toulouse Capitole
Professeur des universités, Université Toulouse Capitole

AUTRES MEMBRES DU JURY :
PELÉ Marie (co-encadrante)
NAEGEL Benoît
SARAUX Claire

Chargée de recherche, Université Catholique de Lille
Professeur des universités, Université de Strasbourg
Chargée de recherche, CNRS, Strasbourg

INVITÉ :
PANSANEL Jérôme

Ingénieur de recherche, CNRS, Strasbourg



ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ
CNRS IPHC UMR 7178

THÈSE

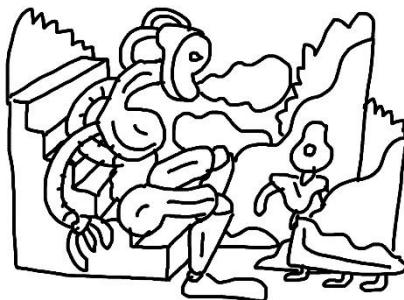
présentée par

Benjamin Beltzung

soutenue le : 17 octobre 2023

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Bioinformatique



Utilisation de réseaux de neurones convolutifs pour mieux comprendre l'évolution et le développement du comportement de dessin chez les Hominidés

THÈSE dirigée par :
SUEUR Cédric

Maître de conférences, Université de Strasbourg

RAPPORTEURS :
SOULÉ-DUPUY Chantal
TESTE Olivier

Professeure des universités, Université Toulouse Capitole
Professeur des universités, Université Toulouse Capitole

AUTRES MEMBRES DU JURY :
PELÉ Marie (co-encadrante)
NAEGEL Benoît
SARAUX Claire

Chargée de recherche, Université Catholique de Lille
Professeur des universités, Université de Strasbourg
Chargée de recherche, CNRS, Strasbourg

INVITÉ :
PANSANEL Jérôme

Ingénieur de recherche, CNRS, Strasbourg

À ma mère.

Remerciements

Tout d'abord, je souhaiterais grandement remercier mes encadrants, Cédric et Marie. Merci à vous de m'avoir offert l'opportunité de continuer mon travail en thèse, à la suite de mon stage de master dans ce fascinant projet sur le comportement de dessin, d'ailleurs très exploratoire, et où vous êtes toujours restés très encourageants. Merci à vous pour votre confiance et l'autonomie que vous m'avez accordées tout au long de ma thèse, et pour tout ce que vous avez pu m'apporter, tant scientifiquement qu'humainement. J'aimerais aussi vous remercier pour votre bienveillance dans des conditions qui n'ont pas toujours été évidentes, que ce soit pendant les périodes de confinement ou non. Malgré la distance, vous avez toujours été présents et réactifs quand j'en avais besoin.

J'aimerais également remercier Lison, sans qui je n'aurais probablement jamais pu participer à ce projet de recherche, et qui m'a avec Cédric et Marie permis de faire mes premiers pas dans le monde de la recherche à travers nos nombreuses et longues discussions sur les fractales.

Merci aux membres du Département d'Écologie, Physiologie et Éthologie de leur accueil, et en tant que novice en biologie, je tiens à les remercier de m'avoir initié à la biologie. En particulier, j'aimerais remercier Anouch, Nicolas, Claire et Vincent, et plus récemment Clément et Cyrielle, pour l'organisation régulière des BEEPS. En plus de l'aspect scientifique, j'aimerais remercier Élisabeth de m'avoir guidé pour toutes les démarches administratives.

Merci à tous les doctorants pour la bonne ambiance de ces trois dernières années. En particulier, j'aimerais remercier Anouch, Nicolas et Hugo avec qui j'ai partagé le bureau tout au long de ma thèse.

Merci également à Julien Renoult pour tes suggestions durant la thèse, qui m'ont permis de mieux appréhender les réseaux de neurones convolutifs et les approches possibles pour les interpréter, ainsi que pour ta collaboration sur nos deux articles.

I would like to thank Shinya Yamamoto and Hirata Satoshi, without whom my stay in Japan would not have been possible and for their hospitality during my stay at the WRC. J'aimerais également remercier Julien Paulet, avec qui j'ai pu partager l'émerveillement de découvrir le Japon.

Je souhaiterais grandement remercier mes amis et ma famille, en particulier mes parents, qui ont toujours été là pour moi et qui m'ont soutenu et encouragé tout au long de ma thèse.

J'aimerais finalement remercier les membres de mon comité de suivi de thèse de m'avoir aiguillé lors de mes deux premières années de thèse, ainsi que les membres du jury de thèse de m'avoir fait l'honneur d'évaluer mon travail et d'y apporter un regard extérieur.

Liste de publications et de présentations orales

Articles présents dans le manuscrit

Beltzung, B., Pelé, M., Sueur, C. Using CNN predictions to study the drawing development. *in prep.*

Beltzung, B., Pelé, M., Renoult, J., Sueur, C. (2023). Deep learning for studying drawing behavior: A review. *Frontiers in Psychology*

***Beltzung, B.**, *Martinet, L., MacIntosh, A., Meyer, X., Hosselet, J., Pelé, M., Sueur, C. (2022) To draw or not to draw: understanding the temporal organization of drawing behaviour using fractal analyses. *Fractals*

Beltzung, B., Pelé, M., Renoult, J., Shimada, M., Sueur, C. (2022). Using Artificial Intelligence to Analyze Non-Human Drawings: A First Step with Orangutan Productions. *Animals*

Sueur, C., Martinet, L., **Beltzung, B.**, Pelé, M. (2022). Making drawings speak through mathematical metrics. *Human Nature*

Sueur, C., **Beltzung, B.**, Pelé, M. (2022). La création artistique non humaine : quel degré d'auteurisation pour les animaux et les machines ? *Polygraphe(s)*

Communications orales

2023. Sketching out the mind: A mathematical and AI approach to understanding drawing behavior. WRC, Kyoto, Japan.

2022. How can artificial intelligence improve our understanding of drawing behavior? Ecology & Behaviour, Strasbourg, France.

2022. Introduction to deep learning. BEEPS seminar. DEPE - IPHC, CNRS, Strasbourg, France.

2021. Interprétation et compréhension du comportement de dessin chez les Hominidés par l'intelligence artificielle. Séminaire Anthro-Lab, Université Catholique de Lille, France.

2020. Workshop imaginecology: Deep learning for image and sound processing and analysis in ecology.

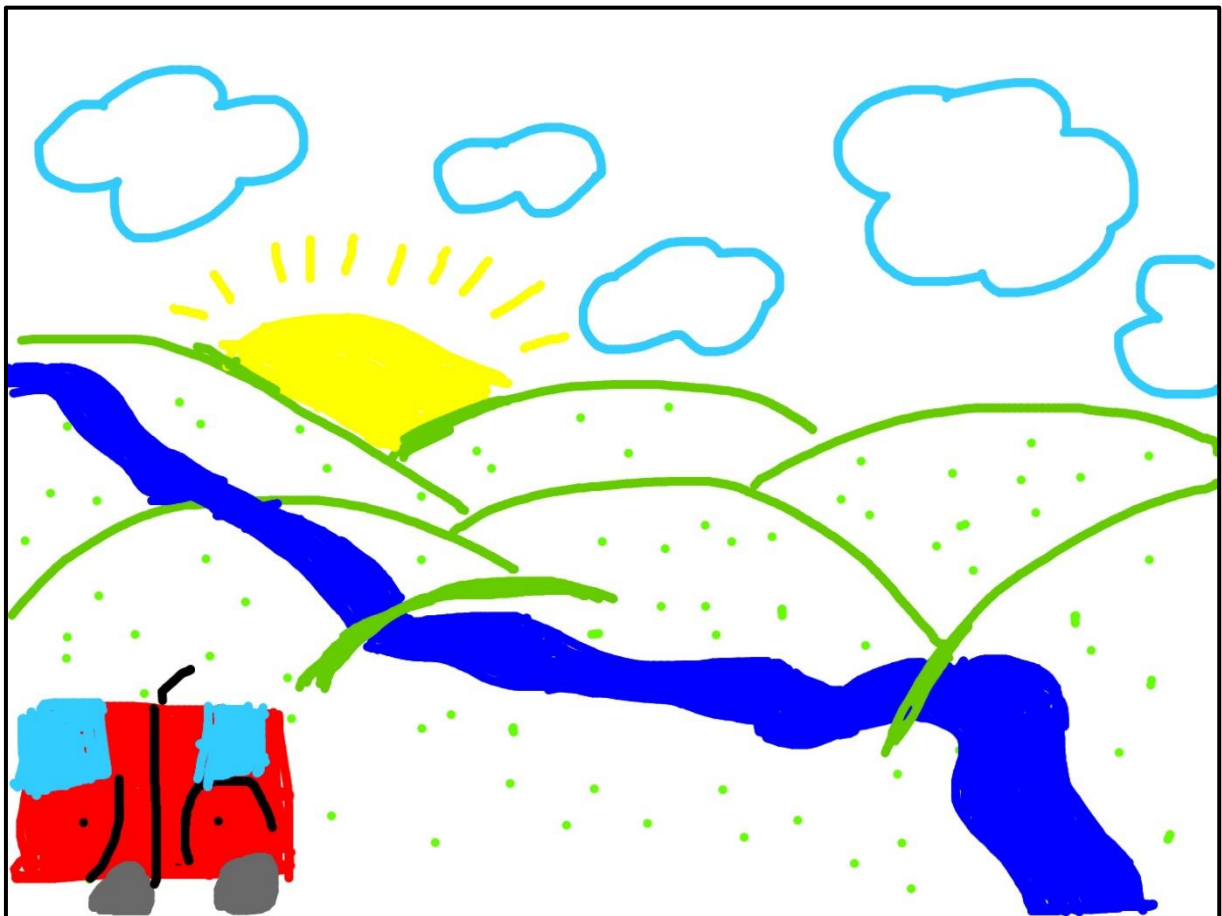
Table des matières

Chapitre 1	Introduction	1
I.	Le dessin	4
1)	Le dessin, c'est quoi?.....	4
a.	Le point de départ du dessin.....	4
b.	Le dessin comme moyen	6
c.	Le dessin comme produit final	8
2)	Le lexique du dessin	10
3)	Origine et développement du comportement de dessin	11
a.	La phylogénie du dessin.....	11
b.	Du gribouillis à l'œuvre d'art	16
II.	S'affranchir de la subjectivité de la lecture d'un dessin	23
1)	Des premiers indices mathématiques... ..	23
2)	... à l'intelligence artificielle	27
a.	Bref historique.....	27
i)	L'intelligence artificielle : de la mythologie à l'informatique	27
ii)	Machine Learning.....	29
iii)	Deep Learning	31
b.	Théorie des CNNs	33
i)	Architecture des modèles.....	33
ii)	Entraînement des modèles	39
c.	Amélioration des performances d'un CNN	42
III.	Problématique	47
1)	Objectif	47
2)	Article 1 : Et le deep learning appliqué au dessin?.....	51
Chapitre 2	Une première approche mathématique	99

I. Article 2 : La complexité temporelle du dessin	101
II. Article 3 : Fusion d'indices mathématiques pour déchiffrer le dessin.....	137
Chapitre 3 Chez l'humain : du gribouillage à la maîtrise	183
I. Article 4 : Quantifier le développement de dessin	185
Chapitre 4 L'intentionnalité chez les non-humains	209
I. Article 5 : Analyse exploratoire de dessins d'orang-outan	211
II. Article 6 : Les animaux et les machines sont-ils des artistes?.....	239
Chapitre 5 Discussion.....	259
I. Limites et perspectives du deep learning	260
1) Que voit un modèle de deep learning?	261
a. CNN et système visuel humain	261
b. Texture et forme	262
2) Interprétabilité du deep learning.....	264
a. Interprétation visuelle de réseaux neuronaux.....	264
b. Détection de concept	265
c. Approche non supervisée	266
3) Des bases de données plus complètes.....	268
II. Conclusion	270
Références.....	273
Annexes	283
I. Annexes de l'Article 2	283
II. Annexes de l'Article 3	288

Chapitre 1

Introduction



Le dessin est un comportement présent chez l'être humain depuis des millénaires, mais présente toujours de nombreux mystères. Ce comportement est-il propre à l'être humain ? Pourquoi avons-nous un tel attrait pour le dessin, dès notre plus jeune âge ? Les gribouillis des jeunes enfants ont-ils une signification ? Afin de mieux comprendre ce comportement, il est nécessaire de l'analyser ontogénétiquement mais aussi phylogénétiquement, chez nos plus proches voisins encore vivants, les primates non humains. Au départ anecdotiques, des comportements proches du dessin humain ont été observés chez des primates, amenant les chercheurs à imaginer et mettre en œuvre des approches plus méthodologiques pour analyser leurs productions (Martinet & Pelé, 2021). Bien qu'aucune figuration n'ait été observée dans ces dernières, l'intérêt que certains primates portent au dessin demeure fascinant et intrigant, soulevant ainsi de nouvelles questions.

Mais l'origine évolutive du dessin n'est pas le seul aspect de ce comportement dont beaucoup de questions restent sans réponse. Depuis plus d'un siècle, les chercheurs s'intéressent au dessin enfantin afin de comprendre l'ontogénie de ce comportement. Se concentrant d'abord majoritairement sur les formes présentes dans les dessins (Rose et al., 2006), de nombreuses approches ont été considérées depuis.

La complexité du dessin et les mécanismes qui y sont liés permettent de le voir tel un langage (Baldy, 2011 ; Cohn, 2012). Cette richesse rend difficile une compréhension intégrale du dessin. Pour cette raison, ce travail de thèse s'inscrit dans un projet de recherche visant à proposer une approche nouvelle du dessin à travers des indices mathématiques. Plus précisément, ce travail de thèse a pour but de déchiffrer des dessins produits par des humains et orangs-outans, à travers des

indices fractals et des réseaux de neurones convolutifs, qui permettent une analyse plus poussée des caractéristiques présentes dans les dessins.

En premier lieu, cette introduction définit le dessin ainsi que les approches traditionnelles, considérant le dessin selon le *quoi* (les marques laissées sur le support) et le *comment* (le processus permettant d'y arriver) (Vinter, 1999). Les grandes lignes de l'étude du comportement de dessin chez les primates non humains sont ensuite expliquées, ainsi que l'étude du développement du dessin chez l'enfant. L'approche mathématique de cette thèse est détaillée dans la deuxième partie de l'introduction, qui commence par préciser certains indices mathématiques ayant déjà été étudiés dans ce projet de recherche. Une esquisse de l'histoire de l'intelligence artificielle, du machine learning et du deep learning est ensuite expliquée. L'architecture des modèles de deep learning ainsi que les concepts fondamentaux sont alors détaillés. La problématique est ensuite posée, et cette introduction se termine finalement par une revue de la littérature de l'utilisation du deep learning à l'analyse du comportement de dessin.

I. Le dessin

1) Le dessin, c'est quoi ?

Avant de s'intéresser au dessin d'un point de vue comportemental, il est nécessaire de définir ce terme et les implications qui en découlent.

a. Le point de départ du dessin

Caricature, dessin satirique, croquis... La polyvalence du dessin et ses nombreuses utilisations reflètent sa capacité unique à communiquer des idées, des émotions et des messages de manière visuelle. De nature flexible, il peut être produit sur différents supports, à condition qu'une surface forme un arrière-plan (Grottanelli,

1961 ; Pinto & Smith, 1999). Mais, en raison de sa nature extrêmement vaste et diversifiée, en donner une définition consensuelle s'avère difficile. Afin d'avoir une première idée de ce qu'est le dessin, nous pouvons revenir à l'origine de ce mot.

Dérivé de l'italien *disegno*, signifiant à la fois dessin et projet, le terme dessin s'écrivait jusqu'au XVIII^e siècle *dessein*. C'est en 1727 que le lexicographe Furetière le définit alors sous différents angles, à savoir le *dessin*, tel que nous le connaissons, et le *dessein* en tant que « projet, entreprise, intention ». Il le définit également comme « la pensée qu'on a dans l'imagination de l'ordre, de la distribution et de la construction d'un tableau, d'un poème, d'un livre, d'un bâtiment », que nous pouvons notamment voir comme le dessin technique, utilisé en architecture. Le *dessein* à cette époque caractérise donc à la fois le projet et le *dessin* en lui-même. Ce n'est qu'au milieu du XVIII^e siècle que la scission entre les deux termes se produit, laissant ainsi au *dessin* le sens de représentation graphique. Pour autant, ces nuances et difficultés ne sont pas propres à la langue française. En effet, le terme anglais *drawing* peut signifier à la fois un objet concret, une activité, un participe, un gérondif (Ashwin, 2016), ou encore un art. L'objet concret sera ici appelé produit final, tandis que l'activité désignera le processus permettant d'arriver au produit final.

Afin de donner une première définition au dessin tel que nous le connaissons, nous pouvons nous appuyer sur celle de l'artiste (Alcopley, 1968) qui le définit comme une image dans laquelle la composition des lignes prédomine par rapport à la considération des couleurs. Cette définition, relativement simple et purement descriptive, permet alors facilement de distinguer ce qui est un dessin ou non. Seulement, cette définition reste une tentative en ce sens qu'elle présente un défaut majeur. En effet, (Elderfield, 1971), alors étudiant en histoire de l'art (et qui

deviendra plus tard conservateur en chef chargé de la peinture et de la sculpture au Museum of Modern Art à New York) souligne que la définition d'Alcopley exclurait certains dessins et inclurait certaines sculptures ou peintures. Pour Elderfield, deux possibilités s'offrent alors : considérer le produit fini du dessin, en définissant ses attributs et fonctions, ou le considérer comme un processus.

b. Le dessin comme moyen

Bien que de nombreux chercheurs se soient heurtés à la tâche, aucune définition du dessin n'a fait de réel consensus. Toutes ces définitions reposent sur un nombre important de notions de natures hétérogènes. Avant de chercher un consensus sur la définition du dessin, il faut a fortiori considérer les différents domaines dans lesquels il s'applique. En effet, le dessin en architecture n'aura pas la même signification qu'en art. En art contemporain par exemple, bien que de nombreux débats existent sur ce qu'est un dessin, de nombreuses définitions s'accordent sur l'importance de la notion d'intentionnalité (MacDonald, 2014), qui, nous le verrons, est pertinente dans le dessin en général.

Le dessin étant traditionnellement vu comme un moyen (Gombrich et al., 1973) rend presque nécessaire la prise en compte du processus et du dessinateur. Ainsi, pour certains, le dessin est une matérialisation de l'esprit du dessinateur, pour qui c'est par le dessin qu'un sujet peut donner des formes concrètes à ses représentations imaginaires. Suivant cette idée, Grottanelli (1961) propose la définition suivante : « dessiner revient à rendre extrinsèque une image subjective, lui donner une limite, une position dans l'espace, et une forme. »¹ C'est par l'usage de signes et symboles communs que le subjectif artistique peut sortir de son intériorité : une passerelle est

¹ Texte original : « To draw means to make a subjective image extrinsic, to give it a limit, a position in space, and a form. » (Grottanelli, 1961, 460)

formée entre son esprit subjectif et celui du spectateur. Bien plus, ses idées, son sens, ses interprétations sont dorénavant accessibles à autrui, même si l'image est extrinsèquement reliée à l'expérience du dessinateur (Lam, 2009).

Ces idées de matérialisation et de compréhension par un observateur externe sont également présentes dans d'autres définitions, comme une proposée par l'historien de l'art (Mendelowitz, 1967), qui définit le dessin comme « symbole graphique pouvant être lu, car sa signification est communément comprise »², reprenant l'idée de compréhension donnée par Grottanelli.

Cette notion de compréhension et d'interprétation interroge plus encore que le lien entre le dessinateur et son dessin, et remet en question la nature même du dessin. Nous savons que certains animaux sont capables de faire preuve de comportements graphiques (Watanabe, 2012), comme le font les araignées en tissant leurs toiles (Krink & Vollrath, 1997) ou encore les mâles poissons-ballons en formant des rosaces dans le sable pour séduire les femelles (Kawase et al., 2013). En considérant le dessin selon la définition de Mendelowitz, qu'en serait-il alors des productions des animaux non humains ? Si les toiles d'araignées et les rosaces des poissons-ballons ne sont pas visuellement compréhensibles par les humains, elles peuvent avoir un sens pour les autres individus de ces espèces. Ne devrions-nous pas considérer ces marques comme un dessin ? Pour Dexter (2005), « dessiner, c'est être humain »³. Cette assertion forte pousse le dessin au cœur de la nature de l'être humain, qui serait humain parce qu'il dessine. Les marques qu'il laisse dans son environnement à travers le dessin seraient ainsi ce qui le différencie des autres animaux (MacDonald, 2014). Savoir si le dessin

² Texte original : « the drawing [...] is a graphic symbol which can be read because the meaning is commonly understood » (Mendelowitz, Drawing (New York: Holt, Rinehart & Winston, 1967) p. 11.)

³ Texte original : « To draw is to be human » (Dexter, 2005, 6)

doit être considéré comme exclusivement humain ou non fait toujours débat. En effet, de telles questions partagent encore aujourd'hui les chercheurs, qui considèrent pour certains le dessin comme uniquement humain (Ramm, 2005), tandis que d'autres incluent les animaux plus généralement (MacDonald, 2014).

Mendelowitz propose une autre définition du dessin en tant que « rapport de ce qui est vu »⁴. Cette définition, bien éloignée des précédentes, apporte au dessin une notion de représentation. Le but du dessin serait alors de représenter les formes, trop complexes à première vue, du monde qui nous entoure, reflétant ainsi notre perception de ce dernier (Bailey, 1982), et devenant un acte de clarification (Elderfield, 1971). Faut-il alors exclure totalement l'art abstrait du spectre du dessin ? Plus encore que le dessin, la notion de représentation serait alors aussi à définir. Un dessin jugé comme abstrait, mais produit avec une intention de représenter ce qui est vu, serait-il alors considéré comme un dessin ?

c. Le dessin comme produit final

Mais, le dessin, au lieu d'être intrinsèquement défini à travers le lien avec son environnement et son dessinateur, peut être défini de manière bien plus descriptive et graphique. Selon le site du CNRTL⁵, le dessin est une « composition artistique exécutée au crayon, à la plume ou au pinceau. » Le Cambridge Dictionary le définit comme « une image faite avec un stylo, crayon, etc. »⁶. Ces définitions mettent l'accent sur l'outil utilisé pour produire le dessin, plutôt que sa représentation ou son intention. Il est intéressant de noter que le CNRTL ne considère que le crayon, la plume et le pinceau, tandis que le Cambridge Dictionary semble inclure n'importe

⁴ Texte original : « the drawing [...] is a record of what is seen » (Elderfield, 1971)

⁵ <https://www.cnrtl.fr/definition/dessin>

⁶ Texte original : « a picture made with a pencil, crayon etc »
<https://dictionary.cambridge.org/dictionary/english-french/drawing>

quel outil capable de produire des marques (« etc. »). La question du matériel se pose alors. Selon l'Oxford Advanced Learner's Dictionary, le dessin est « une image faite en utilisant un crayon ou un stylo plutôt que de la peinture »⁷. Ici, une différence explicite entre le dessin et la peinture est faite, à travers l'outil utilisé pour produire ces marques.

En plus de l'outil utilisé pour dessiner, certaines définitions mettent l'emphase sur le support. L'*Encyclopaedia Britannica*, dans sa définition du dessin, le propose comme « l'art ou la technique de produire des images sur une surface, généralement du papier »⁸. Ici, la surface proposée est le papier, mais ne semble pas exclure d'autre support.

Mais toutes les marques devraient-elles être considérées comme du dessin? L'intention et l'origine du dessin, considérées comme telles, sont alors importantes. Devrions-nous alors voir les stries glaciaires, comme des dessins, au même titre que des gribouillages produits par des enfants? Et même sans aller si loin, l'écriture fait-elle partie intégrante du dessin? Finalement, la définition donnée au dessin, en tant que produit ou processus, dépendra de la problématique posée.

De plus, au lieu de considérer uniquement le processus ou le produit du dessin, il est important d'avoir à l'esprit que ces deux approches ne sont pas en opposition, mais coexistent et sont complémentaires. En effet, différents processus peuvent mener à des dessins jugés similaires, et un même processus peut conduire à des dessins différents (Freeman, 1972).

⁷ Texte original : « a picture made using a pencil or pen rather than paint »
<https://www.oxfordlearnersdictionaries.com/definition/english/drawing?q=drawing>

⁸ Texte original : « the art or technique of producing images on a surface, usually paper »
(Encyclopaedia Britannica 1991: 4.216)

2) Le lexique du dessin

Qu'il soit défini en tant que processus ou produit, le dessin reste fondamentalement lié à différentes notions, dont les définitions sont alors essentielles. Pour cette raison, (Luquet, 1927), référence incontournable du dessin enfantin, a défini rigoureusement de nombreux termes liés au dessin. Les définitions énoncées ici le sont par souci de clarté, afin de mieux cerner les différentes problématiques liées au dessin. Premièrement, la notion de *représentation* est récurrente dans les tentatives de définition du dessin. Mais qu'est alors la représentation ? Selon le site du CNRTL, la représentation peut être une « reproduction, restitution des traits fondamentaux de quelque chose ou de quelqu'un »⁹, ainsi étroitement liée au produit du dessin plutôt qu'au processus. Le site du CNRTL propose également de définir la représentation comme « ce qui est présent à l'esprit ; ce que l'on "se représente" ». Cette définition, plutôt que liée au produit final, imagine la représentation comme un processus mental, qui serait dans le cas présent associé au processus de création du dessin. Si nous considérons la première définition du CNRTL, comment alors estimer si un dessin restitue les traits fondamentaux ou non de ce qu'il représente ? Afin de répondre à cette question, deux situations sont à considérer. Dans le cas où le dessinateur considère que sa production restitue les traits fondamentaux de ce qu'il voulait dessiner, nous parlons de *représentativité interne*. Si le dessin est également lisible par un observateur externe, correctement ou non, nous parlons alors de *représentativité externe* (Martinet et al., 2021). Dans ce cas, il n'y a aucun doute que cette production a été réalisée avec une intention sous-jacente. Mais qu'est alors l'intention ? En prenant le cas du dessin enfantin, pour Luquet, le dessin n'a pas de finalité et n'est qu'un amusement. Il distingue ainsi l'*intention de*

⁹ <https://www.cnrtl.fr/definition/representation>

l'interprétation : cette dernière est alors « l'idée dans le cours du processus de dessin. C'est la ressemblance qui apparaît à l'enfant après coup, ou en cours d'exécution qui détermine cet écart entre l'intention et l'interprétation » (Jumel, 2015).

Une dernière notion importante quant à la représentation est le lien entre signifié et signifiant (DeLoache, 1991). Ces concepts, fondamentaux en linguistique, le sont tout autant pour un moyen de communication comme le dessin. Prenons l'exemple du mot « arbre ». Lorsque nous entendons ce mot, des images et des idées nous viennent en tête, liées à notre expérience avec ce mot. Ces idées sont ce que nous appelons le *signifié*. Le signifié est l'idée, le concept ou le sens sous-jacent à un ou plusieurs signes. L'ensemble de ces signes est appelé *signifiant*, et est dans ce cas, l'ensemble des phonèmes émis en prononçant « arbre ». Le signifiant permet d'exprimer, de véhiculer l'idée du signifié. Pour le dessin, le signifiant sera le produit graphique du dessin. Il est intéressant de remarquer que dans des langages parlés, comme le français, il peut être extrêmement simple de communiquer des signifiés. En effet, pour exprimer l'idée d'un arbre, il suffit de prononcer les phonèmes du mot « arbre ». En revanche, pour le dessin, bien que des conventions et codes existent également pour représenter l'idée d'un arbre (où les signes seront par exemple deux traits verticaux pour le tronc, et des formes arrondies pour les feuilles), la communication de certaines idées peut être fortement complexe.

3) Origine et développement du comportement de dessin

a. La phylogénie du dessin

Le dessin, comportement omniprésent chez *Homo sapiens* depuis au moins plusieurs dizaines de milliers d'années, présente encore beaucoup de mystères et de questions sans réponse. Les plus anciennes marques découvertes à ce jour remontent à plus de

73 000 ans et ont été produites à l'aide d'un outil de couleur ocre (Henshilwood et al., 2018). Un grand nombre de ces marques semblent avoir été faites avec une intention particulière. Certaines représentant un motif hachuré, d'autres représentant par exemple des lignes sinueuses. Ces lignes, qui peuvent être vues comme basiques, avaient-elles un sens particulier ? Y avait-il une intentionnalité de figuration ? Bien que ces questions soient sans réponse, il n'est pas impossible que l'individu ayant produit ces marques eût un objectif de représentation symbolique. Les plus anciennes figurations remontent quant à elles au Paléolithique. L'une des découvertes les plus célèbres provient de la grotte Chauvet. Datées de plus de 36 000 ans, les peintures murales de cette grotte représentent des animaux, tels que des mammouths, des rhinocéros, des ours, ou encore des bouquetins, majoritairement représentés de profil. La plus ancienne représentation figurative découverte est datée d'au moins 45 500 ans et représente un cochon. Cette peinture a été découverte dans la grotte de Leang Tedongnge, en Indonésie (Brumm et al., 2021). Il est vraisemblable que de plus anciens dessins figuratifs aient existé, mais n'aient pas été découverts à ce jour.

La genèse du dessin et les mécanismes cognitifs associés chez *Homo sapiens* sont donc complexes à étudier. D'autres espèces du genre *Homo* ont également fait preuve de comportement symbolique (García-Diez & Ochoa, 2020), et, bien que sujet à débat, un coquillage gravé et daté de 540 000 ans est attribué à *Homo erectus*, remettant ainsi en question l'origine du dessin et l'exclusivité de ce comportement attribué à *Homo sapiens* (Joordens et al., 2015 ; Martinet & Pelé, 2021). Afin d'étudier plus en détail ces questions, les chercheurs se sont inévitablement tournés vers des espèces proches de la nôtre. Le chimpanzé (*Pan troglodytes*), avec qui nous avons en commun plus de 95 à 98 % de nos gènes (Waterson et al., 2005), est alors

un candidat idéal. En effet, en plus de cette part d'ADN partagée, chimpanzés et humains ont de nombreuses similarités, telles que la fabrication et l'utilisation d'outils (Boesch & Boesch, 1990 ; Shimada & Leys, 2021). Les chimpanzés utilisent par exemple des feuilles pour s'essuyer s'ils se salissent, des bâtons pour ramasser des fourmis ou pêcher des termites, ou encore des pierres pour casser des noix (Beyries & Joulian, 1990). L'utilisation de tels outils, ainsi que la présence de pouces opposables, permet d'imaginer des chimpanzés dessiner comme nous le faisons, à l'aide de crayons ou pinceaux.

Bien qu'aucun chimpanzé n'ait été observé en train de dessiner dans la nature, différents rapports anecdotiques ont démontré l'intérêt de certains chimpanzés pour le dessin (Figure 1) (Kohts, 1935 ; Martinet & Pelé, 2021). En plus d'approfondir nos connaissances sur certains aspects psychologiques de l'apparition du dessin, le dessin chez le chimpanzé permet d'analyser leur organisation perceptuelle (Smith, 1973). En analysant plus de 200 dessins d'Alpha, une femelle chimpanzé, Schiller (1951) démontre que cette dernière est capable de terminer des figures géométriques incomplètes, en accord avec la loi de la clôture issue de la théorie de la Gestalt, selon laquelle le tout est plus que la somme des parties. Suite à cette expérience, Smith (1973) utilise des dessins de trois jeunes chimpanzés captifs dans les mêmes conditions que Schiller 22 ans plus tôt, et ne trouve pas de preuve de la présence des lois de la Gestalt dans les dessins de chimpanzés. Ces résultats, en contradiction avec la précédente étude de Schiller, démontrent la délicatesse de l'interprétation de telles productions, due à l'incapacité des chimpanzés à nous communiquer leurs réelles intentions ainsi qu'à notre tendance naturelle à l'anthropomorphisme. Aucune figuration n'a été à ce jour décelée dans des dessins de chimpanzés ou d'autres grands singes non humains.

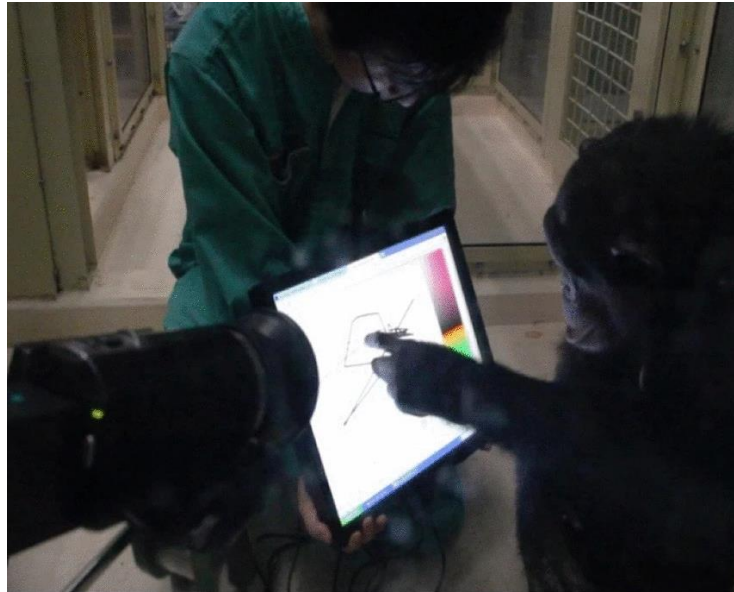


Figure 1. Femelle chimpanzé dessinant au doigt sur écran tactile, issu de Martinet et al. (2021).

Depuis, un nombre non négligeable de chimpanzés en captivité ont produit des dessins, une fois le matériel fourni. Ces derniers utilisaient pour cela leurs doigts ou manipulaient des outils tels que des marqueurs ou des pinceaux, sur un support pouvant être du papier ou encore une tablette tactile. Les dessins produits sur tablette tactile permettent, en plus de récolter les informations spatiales du dessin, de récolter les données temporelles qui y sont liées. Cependant, des primates dessinant sur papier auraient-ils le même comportement en dessinant sur tablette tactile ? Martinet et al. (2023) ont démontré que le comportement de dessin ne variait pas selon le support utilisé chez deux chimpanzés, Ai et Pan. Contrairement au support, les instruments utilisés pour dessiner (une baguette courte, une baguette longue, une barre en métal, un bâton, ou encore leurs doigts) ont un impact sur la complexité des lignes tracées.

Certains chimpanzés en captivité ont dessiné, à des fins d'enrichissement ou d'expérimentation, sans même récompense à la clé (Ladygina-Kots, 2000 ; Morris,

1962 ; Schiller, 1951 ; Smith, 1973). Un tel engouement pour le dessin soulève de nombreuses questions : pourquoi ces individus aiment-ils dessiner ? Quel plaisir tirent-ils de cette exploration ? Est-il purement kinesthésique, ou plus complexe, lié à une réelle intentionnalité quant au résultat final du dessin ?

Bien que les réponses à ces questions ne soient pas connues, le dessin chez les primates non humains continue d'attiser la curiosité des chercheurs, tant chez les chimpanzés que chez d'autres primates, tels que le gorille (*Gorilla gorilla*) (Zeller, 2007), le macaque rhésus (*Macaca mulatta*) (Brewster & Siegel, 1976), ou encore l'orang-outan (*Pongo pygmaeus*) (Hanazuka et al., 2019). Notamment pour ces raisons, le dessin est depuis devenu une méthode courante d'enrichissement pour les primates en captivité. Bien qu'il puisse être surprenant de voir un chimpanzé, un gorille ou encore un orang-outan dessiner, chacune de ces espèces détient des spécificités et doit être considérée avec attention. Les orangs-outans par exemple, peuvent comme les chimpanzés manipuler des outils (Bardo et al., 2017 ; Lethmate, 1982 ; Mendes et al., 2018) et développer des capacités cognitives importantes (Dufour et al., 2009). Particulièrement, et contrairement aux chimpanzés, les orangs-outans sont dotés d'une grande flexibilité au niveau de leurs poignets (Lewis, 1969), permettant plus de possibilités lors de la création de dessins que pour les chimpanzés.

Plus encore que les espèces, il est fondamental de considérer chaque primate dans son individualité à travers sa personnalité, et ne pas tirer de généralisation hâtive sur une espèce en étudiant un nombre limité d'individus (Martinet et Pelé, 2021). Certains individus ne montreront pas d'intérêt pour le dessin, tandis que d'autres seront beaucoup plus stimulés et connaîtront même une évolution de leurs marques (Boysen et al., 1987 ; Kohts, 1935 ; Martinet et al., 2023 ; Morris, 1962). Pelé et al.

(2021) ont démontré, en étudiant des dessins de différents orangs-outans, qu'une différence interindividuelle existait, en considérant le style de leurs dessins (les couleurs utilisées, l'espace rempli, et les formes dessinées). De plus, cette étude a démontré que le comportement de dessin d'une femelle orang-outan nommée Molly, particulièrement stimulée par le dessin, évoluait au fil du temps et selon les saisons. Il est ainsi important de considérer la personnalité de chaque individu.

En plus des particularités de chaque espèce et de chaque individu, l'environnement dans lequel évoluent ces derniers est également à prendre en compte. Des événements extérieurs tels que des naissances ou encore une visite scolaire peuvent impacter la motivation de l'individu (Hanazuka, 2019). Contrairement au cas des primates non humains, une part de cette variabilité peut être contrôlée pour des analyses concentrées cette fois sur l'humain, tant ontogéniques qu'artistiques.

b. Du gribouillis à l'œuvre d'art

Réalisés par des artistes reconnus ou de jeunes enfants, les dessins abstraits ont toujours fasciné et attisé la curiosité d'un œil extérieur. Seulement, comment caractériser la différence entre une œuvre de Jackson Pollock et un gribouillage enfantin ? Le simple gribouillage, sans intention, existe-t-il ?

Pour répondre à ces questions, il est essentiel d'analyser l'émergence et le développement du comportement de dessin chez l'enfant. Dès leur plus jeune âge, les enfants explorent le monde qui les entoure, en utilisant leurs sens, et cherchent à communiquer leurs émotions. Trop jeunes pour les exprimer à travers le langage, le dessin, d'abord sous forme de gribouillis, devient un moyen privilégié pour laisser parler leur créativité et transmettre leurs émotions. Le gribouillis consiste à exécuter « des mouvements de la main qui, celle-ci étant munie d'accessoires variés, laissent sur un support tel qu'une feuille de papier des traces visibles qui n'y étaient pas

auparavant » (Luquet, 1927, p. 109). Il est important de noter que pour Luquet, contrairement au dessin, le gribouillis ne requiert pas d'intention de représenter.

Au fil de sa pratique, les compétences artistiques de l'enfant s'améliorent, et ce moyen de communication devient ainsi un langage visuel à part entière. Le dessin joue alors un rôle essentiel dans le développement cognitif de l'enfant, permettant également d'avoir un aperçu de son état mental à travers la richesse des informations transmises par son dessin. Cette abondance d'informations a stimulé de nombreux chercheurs, permettant au comportement de dessin de devenir un champ de recherche fondamental quant au développement cognitif de l'enfant (Abdulhameed & Rashid, 2022 ; Chappell & Steitz, 1993 ; Hagood, 2003 ; Martinet & Pelé, 2021). Pendant longtemps, la principale assomption faite sur le développement du dessin consistait en un désir inné de l'enfant de produire des dessins réalistes, qui, par manque de capacités cognitives, échouait à la tâche (Luquet, 1913, 1927). Par la suite, il a été démontré que les préférences des enfants en termes de dessins n'étaient pas nécessairement les plus réalistes, et que l'aspect esthétique pouvait entrer en jeu (Jolley, 2009).

En 1927, Luquet distingue, en étudiant les dessins réalisés par sa fille, quatre stades du développement du comportement de dessin. Aux alentours de 3 ans, les premières marques réalisées par l'enfant sont désordonnées et sans intention, purement guidées par un plaisir moteur. À force de dessiner, il prend conscience des marques qu'il produit et ces dernières lui rappellent des objets du monde réel. Cette phase, où la notion de réalisme n'est pas intentionnelle lors du processus de création, est appelée *réalisme fortuit*. L'enfant essaye dès lors d'améliorer ses marques afin de produire des dessins plus réalistes.

Il entre alors dans la phase de *réalisme manqué*, caractérisée par une représentativité interne. À ce stade, les dessins de l'enfant sont dirigés par une intention de réalisme. L'enfant, maladroite de par ses capacités motrices limitées, n'arrive cependant pas à représenter correctement ce qu'il désire. En plus de cette limitation, la concentration requise pour réfléchir à son dessin et contrôler son geste implique une discontinuité dans l'attention enfantine qui complique cette tâche. Au fur et à mesure de ses productions, l'enfant s'améliore et entre dans la phase suivante entre 5 et 8 ans.

L'attention et la concentration de l'enfant s'étant développées, l'enfant utilise alors la connaissance qu'il a du monde qui l'entoure afin de représenter la réalité. Il dessine alors ce qu'il sait. Pour cette raison, ce stade est appelé *réalisme intellectuel*. Ses productions sont alors largement compréhensibles et interprétables par un adulte, faisant ainsi preuve de représentativité externe, devenant donc figuratives, malgré la présence de certaines erreurs. Se basant majoritairement sur ses connaissances, une caractéristique omise par l'enfant sera la transparence. L'enfant dessine des objets qui ne seraient pas visibles en réalité (par exemple une maison ainsi que les personnages ou meubles qui sont à l'intérieur, ou encore le corps d'un personnage sous ses habits). De plus, l'enfant ne prête pas attention à la perspective et ses dessins sont caractérisés par du rabattement. Les pieds des animaux sont par exemple représentés de côté, tout comme les roues des voitures. Son expérience le conduit finalement à remarquer ces erreurs et essayer de les corriger. L'achèvement de cette phase caractérise également la fin du graphisme enfantin (Bresler & Thompson, 2002).

Vers l'âge de 10 ans, il entre dans le dernier stade, appelé *réalisme visuel*. Il est alors capable de produire un dessin entier d'une seule perspective, remplaçant alors

le rabattement par une notion de perspective plus poussée. La transparence dont l'enfant fait preuve dans la phase de réalisme intellectuel est remplacée par de l'occlusion. Comme de nombreuses autres théories, celle de Luquet met en lumière une phase où l'enfant dessine ce qu'il sait, suivie d'une phase où il dessine ce qu'il voit (Freeman, 1993 ; Freeman & Janikoun, 1972 ; Piaget & Inhelder, 1956).

Seulement, comment analyser concrètement le développement du dessin chez l'enfant ainsi que les transitions entre ces différentes phases ? Une option proposée par Baldy (2005) est d'étudier l'évolution des dessins de bonshommes qui peuvent être de différents types, comme montré sur la Figure 2 (Baldy, 2005 ; Goodenough, 1926 ; Luquet, 1920 ; Lurçat, 1979).

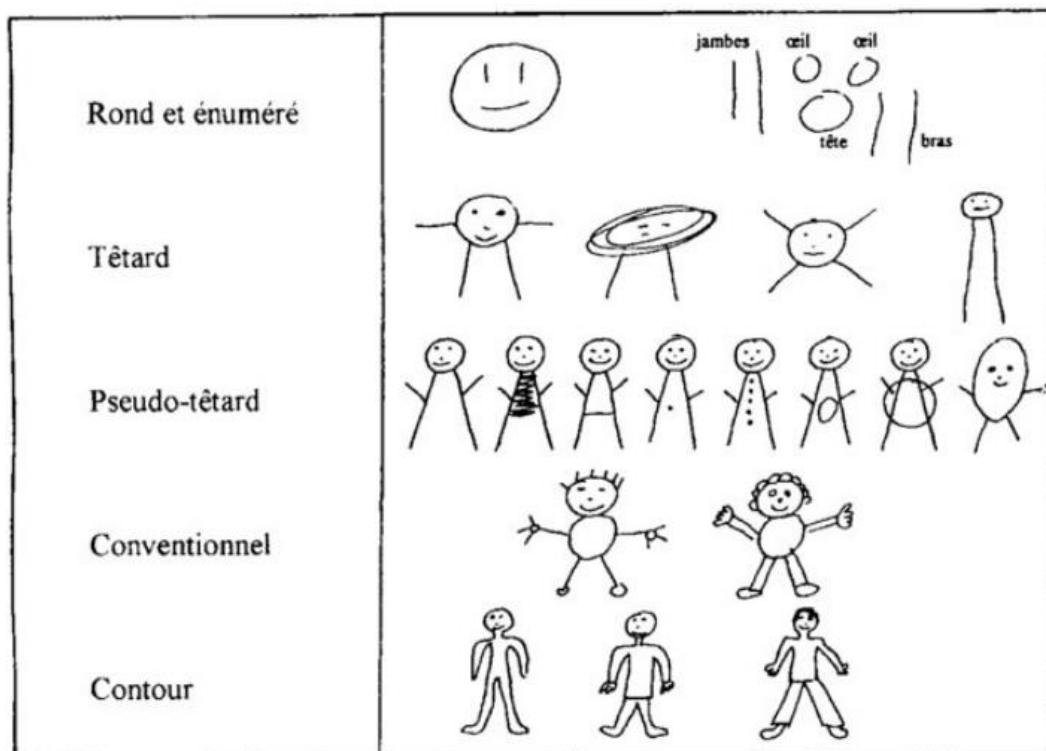


Figure 2. Exemples de dessins pour chaque stade selon la classification de Baldy, issu de Baldy (2009).

Beaucoup de chercheurs, dont Luquet faisait partie, n'étudiaient pas les dessins des plus jeunes enfants. En effet, les prémices du dessin chez les jeunes enfants étaient considérées comme purement kinesthésiques ; guidées par un plaisir moteur ou

perceptif. Cependant, différents travaux ont permis de mettre en lumière des activités préreprésentationnelles dans ces dessins, jugés premièrement sans signification (Cox, 1992 ; Freeman, 1993 ; Gardner & Wolf, 1987 ; Yamagata & Shimizu, 1997). Dunst & Gorman (2009) ont démontré que la durée pendant laquelle un enfant dessine et la complexité du dessin associé sont plus importantes lors de dessins collaboratifs, où l'enfant interagit avec un adulte. De plus, Yamagata (1997) démontre à l'aide d'une étude longitudinale d'une durée de 18 mois sur deux dyades mère-enfant (débutée quand les enfants avaient 12 mois), que dès l'âge de 2 ans, l'enfant peut faire preuve d'intentionnalité dans ses dessins lors d'interactions avec sa mère. Ces différents résultats démontrent que l'émergence du comportement de dessin n'est pas le fruit d'un simple plaisir moteur, mais s'inscrit dans un contexte de communication. Ainsi, un enfant peut être plus ou moins stimulé par le dessin selon l'intérêt que d'autres individus, enfants ou adultes, portent à ses productions, ou encore s'il décide d'imiter les adultes (Quaglia & Saglione, 1976).

Pour toutes ces raisons, et parce que les premières années de vie sont caractérisées par un développement cognitif important, les gribouillis produits par les jeunes enfants ont été observés d'un autre œil. Bien que les premiers dessins, généralement produits vers l'âge de 2 ans, ne paraissent a priori pas marqués d'intentionnalité, cette interprétation ne serait limitée non pas par l'enfant, mais plutôt par l'observateur (Longobardi, 2015). En effet, à cet âge-là, l'enfant est déjà capable de contrôler ses mouvements, par exemple pour manipuler ou attraper des objets (Pinto et al., 2011). Bien avant le dessin figuratif, le processus associé est alors en réalité bien plus complexe que de simples marques aléatoires.

En effet, une des premières indications d'intentionnalité survient lorsque l'enfant produit des *représentations d'action*, également appelées *dessins gestuels*. Ces

productions apparaissent lorsque l'enfant produit des sons, verbalise son dessin. Il utilise alors les marques et le son pour caractériser son dessin, plutôt que la forme de ce dernier. À ce stade, il porte bien plus d'intérêt à l'aspect dynamique des objets qu'à leur forme géométrique (Cox, 1992 ; Freeman, 1993 ; Gardner & Wolf, 1987). À ce stade, le dessin est considéré comme non figuratif, et n'est donc pas interprétable par un observateur externe.

Une fois l'enfant capable de nommer son dessin en exprimant l'objet qu'il représente, il entre dans la phase de *romance*. Bien que capable de nommer sa production, il sera difficile pour un adulte de trouver une ressemblance entre le dessin et l'objet en question. La nomination du dessin peut être faite avant, pendant qu'il dessine, ou une fois la production terminée. Cette phase, caractérisée par l'incapacité d'un adulte à lire correctement le dessin, peut être expliquée par différents facteurs (Adi-Japha et al., 1998). Premièrement, il est possible que l'enfant désigne son dessin par un objet sans qu'il n'ait réellement voulu le représenter, et sans qu'il y trouve réellement de ressemblance (Freeman, 1993 ; Gardner, 2017). Deuxièmement, il est tout simplement possible que ses talents de dessinateur ne permettent pas à un observateur externe la compréhension de sa production, restant alors dans une phase de représentativité interne seulement. Troisièmement, il est possible que, de manière plus marquante que du réalisme fortuit, les marques laissées par l'enfant lui rappellent un objet a posteriori.

L'*élicitation guidée* survient lorsque les productions libres de l'enfant ne laissent paraître aucune intentionnalité pour un observateur externe, mais que l'enfant est capable de produire un dessin figuratif une fois assisté, ou encore de compléter des dessins déjà commencés. La phase d'élicitation guidée est notamment observable

dans les situations énoncées plus tôt, où l'enfant interagit avec d'autres enfants, des adultes, ou encore sa mère.

Les différentes théories existantes quant à l'émergence et le développement du comportement de dessin permettent à la recherche d'avancer, mais certaines difficultés demeurent. L'âge de l'enfant n'est pas le seul facteur à considérer. En effet, certaines variables comme la culture ou encore le niveau d'urbanisation du pays du dessinateur peuvent affecter la représentation et le contenu du dessin (Restoy et al., 2022).

Si le comportement de dessin est un champ actif de la recherche, le point de vue du dessinateur est rarement considéré (Ramm, 2005). Les productions des chimpanzés et orangs-outans sont ainsi vues d'un œil humain, rendant les interprétations et analyses sensibles à de l'anthropomorphisme. L'anthropomorphisme est un biais cognitif caractérisé par une tendance à attribuer des caractères et des comportements humains à des animaux non humains, des objets, ou encore des divinités. Cela peut poser problème, étant donné qu'il est difficile de savoir si un même trait dessiné par un orang-outan et un humain aura le même sens. Ce problème existe également en étudiant le comportement de dessin chez l'enfant. Il est difficile de ne pas porter un regard adultocentrique (Danic et al., 2006), c'est-à-dire basé sur l'idée que l'enfant est un adulte miniature, inachevé, estimant ainsi que la qualité de tels dessins est « moins bonne » que celles des adultes, et jugeant d'une volonté d'y tendre. Pour cette raison, la notion d'enfantisme est de plus en plus considérée dans la recherche enfantine. L'enfantisme¹⁰ apparaît « lorsque les enfants sont

¹⁰ Ici, l'enfantisme est défini de manière positive (comme par exemple le terme féminisme), où l'enfant n'est pas un adulte inachevé mais plutôt un groupe à part entière. Ce terme peut également être utilisé de manière négative (comme le terme racisme) pour désigner les discriminations subies par les enfants (McGillivray, 2022).

considérés de façon enchantée comme étant un groupe culturel distinct, les auteurs faisant alors fi des différences internes à ce groupe et notamment du poids fondamental des variables sociales » (Simon, 2014).

Afin de s'affranchir de ces limites, se doter d'outils objectifs, compréhensibles et interprétables pour analyser de telles productions est une étape essentielle. Pour cela, les mathématiques, à travers des indices tels que celui de Gini ou encore les fractales, et l'intelligence artificielle, permettraient de considérer un nombre plus important de caractéristiques ainsi que d'avoir une approche plus mathématique et objective du dessin.

II. S'affranchir de la subjectivité de la lecture d'un dessin

Si le comportement de dessin a été vu sous différents angles, les multiples biais présents dans l'analyse de ce dernier compliquent son étude et rendent son déchiffrement complexe. Les mathématiques peuvent permettre de minimiser ces biais. En effet, des indices définis mathématiquement seront uniquement liés à une représentation objective du dessin, le considérant comme image numérique (c'est-à-dire un tableau représentant des pixels) sur lequel peuvent être appliquées des formules ne reposant pas sur une interprétation visuelle et humaine. En plus de ces indices mathématiques, une approche liée à l'intelligence artificielle permettrait d'atteindre le même objectif.

1) Des premiers indices mathématiques...

Afin de réduire les biais liés à l'interprétation du dessin, une approche mathématique basée sur des indices quantitatifs, tant spatiaux que temporels, ont été explorés dans ce projet de recherche (Martinet et al., 2021) et dans ce travail de thèse,

notamment dans le Chapitre II. Différents indices dont l'interprétation est intuitive existent, tant en considérant l'aspect spatial (distance du dessin), temporel (durée du dessin, vitesse de tracé), ou colorimétrique (couleur moyenne, nombre de couleurs). En revanche, des indices plus complexes requièrent une définition et une interprétation plus poussées. Certains indices sont introduits ci-dessous de manière non exhaustive afin de donner une idée plus claire de ce que des indices mathématiques peuvent renseigner sur le comportement de dessin.

Fractales spatiales

Réputées pour représenter des structures naturelles, les fractales spatiales, introduites par Mandelbrot (1967), ont été utilisées pour décrire des structures dans différents domaines tels que les neurosciences (Fernández & Jelinek, 2001), la géostatistique (Castrignano & Stelluti, 1999), ou encore la physiologie (Glenny et al., 1991), et permettent de caractériser « l'irrégularité et [...] la complexité des formes naturelles » (Robert & Roy, 1993). Un objet spatial est un objet dont les formes géométriques se répètent, sont similaires, à différentes échelles. En écologie, les fractales spatiales sont utilisées pour décrire les mouvements optimaux des animaux (Sueur, 2011 ; Viswanathan et al., 1999). Pour ces raisons, ces dernières peuvent permettre d'analyser la complexité spatiale des dessins, et plus précisément encore l'efficacité d'un dessin (Martinet et al., 2021). En étudiant des dessins de chimpanzés et d'humains de différents âges, Martinet et al. (2021) ont pu démontrer que les dessins de chimpanzés étaient moins efficaces que ceux d'humains, et que l'efficacité des dessins humains atteignait son maximum entre 5 et 10 ans, avant de diminuer chez les adultes, qui par l'ajout de détails, perdent en efficacité.

Fractales temporelles

En plus de permettre l'étude des formes géométriques, les fractales permettent d'analyser des séries temporelles, et par extension des processus. Comme son nom l'indique, l'échelle étudiée par cette approche est temporelle. Ainsi, l'aspect fractal temporel est caractérisé par la répétition d'un motif (baisse, hausse) à différentes échelles (seconde, heure, jour, semaine, etc.). Les fractales temporelles permettent d'estimer si le processus en question est caractérisé par de l'anti-corrélation (comportement complexe anti-persistant : une hausse de la série sera suivie par une baisse et inversement), de l'aléatoire (pas de dépendance à long terme), ou encore une corrélation à long terme (forme persistante de mémoire longue). Les fractales temporelles sont très utilisées en finance (Alvarez-Ramirez et al., 2008 ; Ding et al., 2021 ; Eom et al., 2008), en physiologie (Eke et al., 2002 ; Glenny et al., 1991), ou encore en neurosciences (Weiss et al., 2009). En écologie, elles sont notamment utilisées pour étudier le comportement de fourragement en analysant les séquences de plongée chez les oiseaux marins. MacIntosh et al. (2013) ont par exemple démontré que les séquences de plongée présentaient une dépendance à long terme persistante. Ainsi, de longues (respectivement courtes) séquences de plongée seront suivies par de (respectivement courtes) longues séquences de plongée. Appliqué au dessin, cet indice permettrait de caractériser la complexité de ce comportement.

Entropie de Shannon

En théorie de l'information, l'entropie de Shannon est un indice mathématique qui peut être interprété comme une mesure de l'incertitude d'une source d'information et de la distribution de probabilité associée (Shannon, 1948). La source d'information est ici une série temporelle. Plus cette série prendra un nombre différent de valeurs

(avec des probabilités similaires), plus l'information contenue y sera importante, et plus l'entropie sera grande. Cet indice est utilisé dans différents domaines tels que l'imagerie médicale (De Araujo et al., 2003), la cryptographie (Cachin, 1997), ou encore la compression de données (Devetak & Winter, 2003). L'idée derrière l'interprétation de l'entropie de Shannon repose sur le même principe que les fractales temporelles : à savoir quantifier la variabilité, la dynamique de la série temporelle à différentes échelles (Ma et al., 2018). Couplée aux indices fractals, l'entropie de Shannon permettrait d'évaluer la quantité d'information présente dans un dessin d'un point de vue temporel.

Indice de Gini

L'indice de Gini est un outil statistique permettant de mesurer l'inégalité de la répartition d'une variable. Très populaire pour estimer le degré d'inégalité des revenus (Chitiga et al., 2014 ; Wan, 2001), la valeur de cet indice permet dans ce cas d'estimer si tous les individus ont la même richesse, ou si un seul individu possède une grande majorité des revenus. En considérant le processus de dessin, il pourrait permettre d'estimer les inégalités (la variabilité) au sein de la série temporelle associée.

Ces indices permettent chacun d'étudier un aspect propre et spécifique du dessin, mais ne permettent pas une interprétation globale de ce dernier. Les considérer conjointement permettrait alors une plus grande compréhension de ces indices, ainsi que des interactions qui les relient.

2) ... à l'intelligence artificielle

a. Bref historique

i) L'intelligence artificielle : de la mythologie à l'informatique

L'idée d'une machine intelligente a toujours été présente dans l'esprit de l'être humain, bien avant d'être imaginée en informatique. Dans la mythologie grecque, Talos, un Géant de bronze, est parfois présenté comme un automate forgé par Héphestos afin de défendre la Crète. En philosophie, Condillac (1754) propose dans son *Traité des sensations* de mener une expérience de pensée en imaginant une statue d'abord totalement vide et sans aucun contact avec l'extérieur. Cette statue se voit ensuite pourvue d'un sens, l'odorat, avant de faire l'expérience de différentes odeurs. Cette expérience de pensée n'est pas sans rappeler l'intelligence artificielle, plus précisément l'apprentissage automatique ou machine learning (ML) en anglais, où un modèle au départ sans connaissance (ici la statue), « découvre » et « apprend » de ses expériences (ici les odeurs). Cependant, la philosophie n'est pas le seul domaine précurseur de l'intelligence artificielle. De nombreux écrivains de science-fiction ont également imaginé des machines intelligentes, tant utopiques, comme l'androïde Data de la franchise *Star Trek*, que dystopiques, comme l'ordinateur de bord HAL 9000 dans *2001, l'Odyssée de l'espace* (Kubrick, 1968) ou encore le monstre de Frankenstein dans *Frankenstein ou le Prométhée moderne* (Shelley, 1818). C'est de cette dernière œuvre qu'est née l'expression d'Isaac Asimov de « complexe de Frankenstein », décrivant la peur des humains qu'une machine se rebelle contre son créateur. Pour contrecarrer cette possibilité, Isaac Asimov définit en 1942, dans sa nouvelle *Cercle vicieux (Runaround)*, les trois lois de la robotique (voir Box 1).

Box 1. Lois de la robotique (Asimov, 1967)

1) Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger.

2) Un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres entrent en contradiction avec la première loi.

3) Un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec la première ou la deuxième loi.

Bien que le concept d'intelligence artificielle soit difficile à dater, un des premiers automates a été créé en 1912 par Leonardo Torres Quevedo. Dénommé *El Ajedrecista*, ce robot, doté d'un bras mécanique, avait la capacité de jouer aux échecs. El Ajedrecista ne pouvait jouer qu'à la condition de n'avoir qu'un roi et une tour contre un roi seul, laissé à un joueur humain, et ne pouvait pas perdre. Ce n'est qu'en 1950 qu'Alan Turing, dans son article *Computing Machinery and Intelligence*, se pose la question suivante : les machines peuvent-elles penser ? Il propose alors un test, d'abord appelé jeu d'imitation puis test de Turing. Ce dernier consiste à faire interagir textuellement et à l'aveugle un humain avec deux entités : un autre humain et un ordinateur. L'individu évalué doit deviner lequel de ses interlocuteurs est un ordinateur. S'il se trompe, la machine a passé le test et est dès lors considérée comme intelligente. Ce test est à ce jour encore une référence en la matière. Un des pionniers de l'intelligence artificielle, John McCarthy, définit six ans plus tard, lors de la conférence de Dartmouth (qui marquera le début de l'intelligence artificielle comme science et discipline à part entière), le terme d'intelligence artificielle comme « la science et l'ingénierie de la fabrication de machines intelligentes, en particulier de programmes informatiques intelligents ».

À ce jour, aucun consensus n'existe sur la définition de l'intelligence artificielle (Bhatnagar et al., 2018 ; Brachman, 2006 ; Hearst & Hirsh, 2000 ; Kirsh, 1991). Certains chercheurs considèrent alors une définition déjà existante, d'autres définissent la leur, tandis que certains considèrent que nous n'avons pas besoin de définir ce qu'est l'intelligence artificielle (Monett et al., 2020 ; Wang, 2019).

ii) Machine Learning

Les nombreuses manières d'imaginer une machine intelligente ont mené à la division de l'intelligence artificielle en différents domaines de recherche (Figure 4). Parmi eux, le ML regroupe l'ensemble des méthodes basées sur l'idée d'une machine « apprenant » des données qu'elle reçoit. Les prémices de ce domaine de recherche remontent à l'article *A logical calculus of the ideas immanent in nervous activity*, où McCulloch & Pitts (1943) modélisent pour la première fois un réseau de neurones. En 1952, Arthur Samuel, un pionnier de l'informatique, développe le premier programme capable de jouer au jeu de dames et de s'améliorer au fur et à mesure des parties. Ce programme est notamment réputé pour avoir battu le 4^e meilleur joueur des États-Unis. Sept ans plus tard, en 1959, il met en avant le terme de machine learning. Ce domaine a depuis connu d'importants progrès techniques et est très utilisé dans différents domaines comme l'éducation (Kucak et al., 2018), la finance (Dixon et al., 2020), ou encore la médecine (Rajkomar et al., 2019). De nombreux algorithmes de ML existent et se divisent en plusieurs catégories, selon leur méthode d'apprentissage (voir Box 2). Bien que les intérêts des algorithmes de ML soient multiples, une de leurs limites est la nécessité d'extraire préalablement des caractéristiques. En effet, les variables utilisées dans ces modèles doivent en premier lieu être collectées, généralement par des intervenants humains. Pour de grands jeux de données, cela peut démultiplier les erreurs présentes dans les

données, et peut être une tâche longue et laborieuse, notamment pour des données non structurées, telles que des images, des vidéos, ou du contenu audio, qui peuvent être riches en information, et en extraire la totalité peut s'avérer complexe.

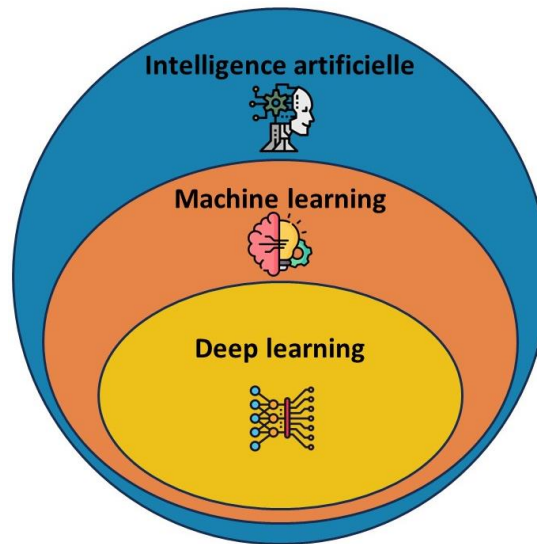


Figure 4. Diagramme de Venn de l'intelligence artificielle.

Pour ces raisons, une sous-branche du ML se concentre sur des modèles permettant de répondre à ces problématiques à travers les réseaux neuronaux : le deep learning, qui excelle depuis plusieurs années dans de nombreux domaines impliquant de l'analyse d'images, de vidéos ou de sons.

Box 2. Types d'apprentissages en machine learning

Il existe différents types d'apprentissage en ML. L'apprentissage dit supervisé consiste à modéliser le lien entre des variables explicatives et une variable réponse. En d'autres termes, utiliser des variables pour prédire les valeurs d'une variable à expliquer. Quand la variable réponse est qualitative, les algorithmes supervisés utilisés sont des algorithmes de classification (Support Vector Machine, régression logistique, classification naïve bayésienne). Si la variable réponse est quantitative, l'apprentissage supervisé permet de prédire les données à l'aide d'une régression (Support Vector Regression, forêt aléatoire).

Il existe ensuite, l'apprentissage non supervisé, qui contrairement à l'apprentissage supervisé, ne requiert pas de variable réponse. L'apprentissage non supervisé permet par exemple du clustering (regrouper les observations les plus similaires entre elles) dont le nombre de groupes peut être défini à l'avance ou non, selon la méthode utilisée. Une autre utilité de l'apprentissage non supervisé réside dans la réduction de dimensions.

iii) Deep Learning

Comme pour le ML, la publication de Warren McCulloch et Walter Pitts (1943) qui proposent le premier modèle mathématique d'un neurone biologique qu'ils appellent neurone formel, est ici fondamentale. Ce modèle est à la base des réseaux de neurones utilisés à ce jour. En 1957, Frank Rosenblatt invente un algorithme d'apprentissage supervisé, le perceptron, un réseau de neurones basé sur le neurone formel inventé 14 ans auparavant. Au milieu des années 1970, les résultats de l'IA n'étant pas jugés assez prometteurs et satisfaisants, le domaine connaît une baisse d'intérêt et de financements. Appelée « hiver de l'IA », cette période durera

jusqu'au début des années 1980. Kunihiko Fukushima (1980) invente alors le neocognitron, le premier réseau neuronal convolutif (CNN), qui deviendra plus tard un modèle fondamental en classification d'images. En 1986, Rina Dechter introduit dans son article *Learning While Searching in Constraint-Satisfaction Problems* l'expression utilisée à ce jour de deep learning (DL), un sous-ensemble du ML regroupant les algorithmes utilisant des réseaux de neurones. Le DL est majoritairement utilisé pour l'analyse de données non structurées, telles que les images, photos, et vidéos. L'IA connaît ensuite un deuxième hiver de 1987 à 1993. Durant cette période, en 1989, Yann LeCun, un des pères fondateurs du DL, propose la première démonstration pratique d'un CNN combiné à de la back-propagation (une méthode d'entraînement de réseaux neuronaux) pour classifier des chiffres écrits à la main. En 2010, une compétition annuelle de détection et classification d'objets dans des images issues de la base de données ImageNet (Deng et al., 2009), nommée ImageNet Large Scale Visual Recognition Challenge (ILSVRC) est lancée. Cette dernière devient une référence en termes d'évaluation d'algorithmes. Ce n'est qu'en 2012 que le secteur du DL explose, quand une architecture de DL, *AlexNet* (Krizhevsky et al., 2012), remporte cette compétition. La longue période entre la création de ce domaine et son explosion est due notamment à la puissance de calcul des ordinateurs, qui était à l'époque trop limitée pour ce type de modèles, ainsi qu'à l'explosion du big data, nécessaire pour entraîner les modèles de DL, rendant obsolètes les méthodes traditionnelles d'analyses. Depuis, le domaine s'est largement démocratisé et influence de manière importante de nombreux secteurs. Tout comme le ML, le DL est très efficace dans de nombreux domaines tels que l'imagerie médicale (Suzuki, 2017) ou encore la prédiction de tremblements de terre (Li et al., 2020).

b. Théorie des CNNs

i) Architecture des modèles

Les CNNs sont des modèles de DL capables de résoudre différentes tâches, telles que de la classification ou encore de la segmentation. Bien que ces modèles puissent prendre en entrée différents formats (image, vidéo, son), nous nous concentrerons ici sur les images. Une fois une image donnée au modèle, de nombreuses couches internes au modèle transforment les données afin d'en extraire des caractéristiques, avant de renvoyer à l'utilisateur une sortie dont les dimensions dépendent de la tâche à résoudre. Le terme de « deep » learning fait référence à la profondeur des réseaux de neurones composés de plusieurs couches cachées. L'architecture et l'apprentissage de tels modèles reposent sur des bases mathématiques et de nombreux concepts propres aux CNNs (Table 1). L'utilisation ainsi que la mise en place de ces derniers dépendent du problème à résoudre, rendant leur compréhension fondamentale.

Tout d'abord, la première couche d'un modèle, appelée la couche d'entrée, permet à un CNN d'analyser les informations des pixels qui seront donnés en entrée. La dimension de cette première couche dépendra notamment de la qualité d'image ainsi que du nombre de canaux (3 pour une image RGB (rouge vert bleu), 1 pour une image en noir et blanc) à considérer.

La force des CNNs pour l'analyse d'image réside dans les couches de convolution qui permettent d'apprendre et détecter des motifs complexes (Figure 5). Une couche de convolution est composée de plusieurs filtres, permettant à chacun de détecter un motif sur l'image de départ. Un filtre est une matrice (généralement de taille 3×3 ou 5×5), au départ aléatoire, puis optimisée lors de l'entraînement du modèle. Le

degré de détection (l'activation) d'un filtre donné sera calculé de la manière suivante, à travers une opération de convolution (Figure 6) :

- Le filtre est superposé sur une région de l'image d'entrée.
- Le produit scalaire (multiplication élément par élément) entre le filtre et la région de l'image est calculé.
- Les éléments de la matrice résultante de ce produit scalaire sont ensuite sommés.
- Le résultat de ce calcul est gardé en mémoire (ainsi que sa position par rapport à l'image).
- Le filtre « glisse » sur l'image d'entrée d'un nombre de pixels défini à l'avance, appelé *pas* (stride).

Le processus est réitéré jusqu'à avoir été appliqué sur toute l'image. Les filtres les moins profonds détectent des caractéristiques simples, telles que des lignes, des couleurs, tandis que les filtres de couches plus profondes détectent des caractéristiques de haut niveau tels que des formes et des objets (Ribeiro et al., 2016).

Le résultat de ce filtrage par convolution est une matrice appelée *feature map* (carte d'attribut), ou encore *activation map* (carte d'activation). Les valeurs de ces matrices représentent le degré d'activation du filtre dans la région correspondante. Plus la caractéristique détectée par le filtre sera présente, plus la valeur dans la matrice sera élevée.

Table 1. Lexique des principaux termes utilisés en DL.

Terme français	Terme anglais	Courte définition
Poids et biais	Weights and biases	Paramètres d'un CNN.
Filtre	Filter, kernel	Une matrice représentant un motif spécifique.
Convolution	Convolution	Application d'un filtre à une image.
Couche de convolution	Convolutional layer	Couche appliquant différents filtres à l'image d'entrée à travers des opérations de convolution.
Pas	Stride	Paramètre contrôlant le nombre de pixels duquel se déplace le filtre à chaque étape lors de l'opération de convolution.
Fonction d'activation	Activation function	Fonction permettant de transformer les données afin d'apprendre des motifs complexes.
Couche entièrement connectée	Fully-connected layer	Couche où chaque neurone est une combinaison linéaire de tous les neurones de la couche précédente.
Carte d'attribut, carte d'activation	Feature map, activation map	Résultat d'une convolution d'une image pour un filtre particulier.
Couche de pooling	Pooling layer	Couche permettant de réduire la taille de la couche précédente en préservant l'information.
Fonction de perte	Loss function	Fonction quantifiant l'écart entre la vraie valeur et sa prédiction.
Ensemble d'entraînement	Training set	Ensemble sur lequel le modèle sera entraîné.
Ensemble de validation	Validation set	Ensemble sur lequel le modèle sera évalué.
Rétropropagation	Back-propagation	Algorithme permettant de calculer le gradient pour chaque poids du modèle.
Optimiseur	Optimizer	Fonction permettant d'ajuster les poids du modèle à l'aide des gradients calculés par rétropropagation.
Taux d'apprentissage	Learning rate	Paramètre quantifiant à quel point les poids sont ajustés.
Taille de lot	Batch size	Nombre d'exemples d'entraînement donnés au modèle avant la mise à jour des poids.
Nombre d'epochs	Number of epochs	Nombre de fois où tout le jeu d'entraînement aura été donné au modèle.
Courbe d'apprentissage	Learning curve	Représentation graphique de l'évolution d'une métrique (perte ou précision) au fur et à mesure de l'entraînement du modèle.
Surapprentissage	Overfitting	Phénomène survenant quand le modèle est trop spécifiquement adapté au jeu d'entraînement et généralise mal.
Apprentissage par transfert	Transfer learning	Utilisation des connaissances acquises par un modèle préalablement entraîné pour une autre tâche.

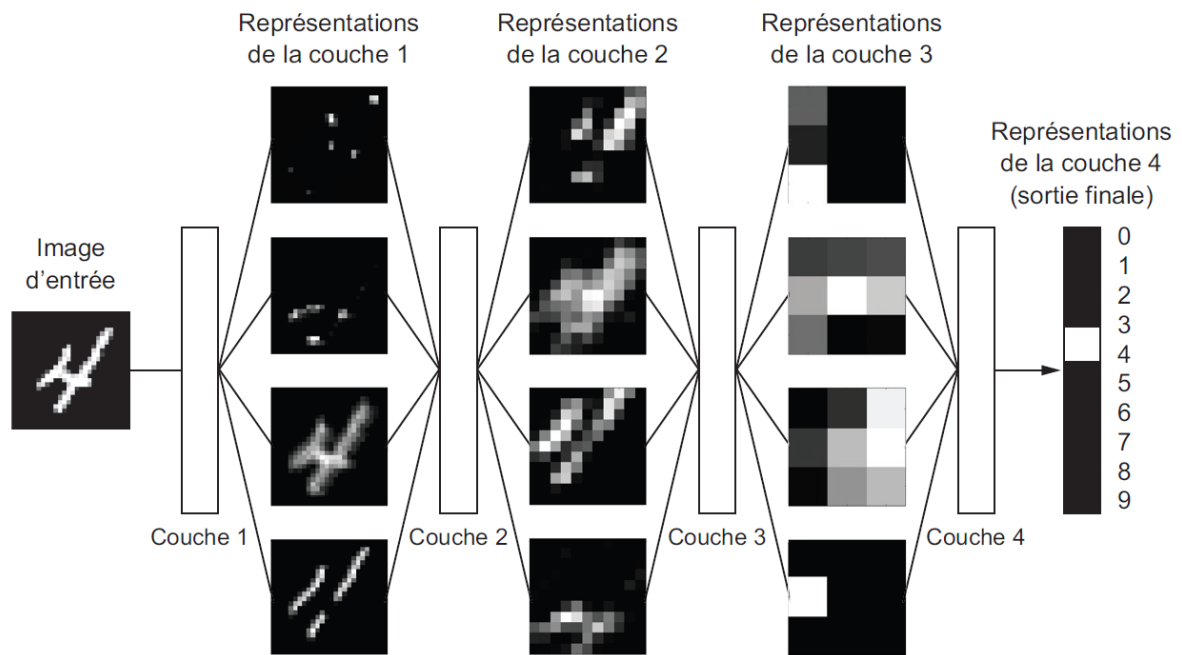


Figure 5. Illustration schématique d'un CNN. Le modèle présenté ici réalise de la classification de chiffre. Pour l'image d'entrée, représentant le chiffre 4, des exemples de cartes d'activation sont montrés pour les différentes couches. En regardant par exemple la carte d'activation la plus en bas pour la couche 1, nous voyons que seules les diagonales droites sont détectées. En d'autres termes, le filtre associé détecte la présence de diagonales droites. Les filtres de la couche suivante détecteront des caractéristiques plus complexes et globales. Ce processus est réitéré jusqu'à la couche de sortie, où le modèle classifie finalement l'image d'entrée, d'après Chollet (2020).

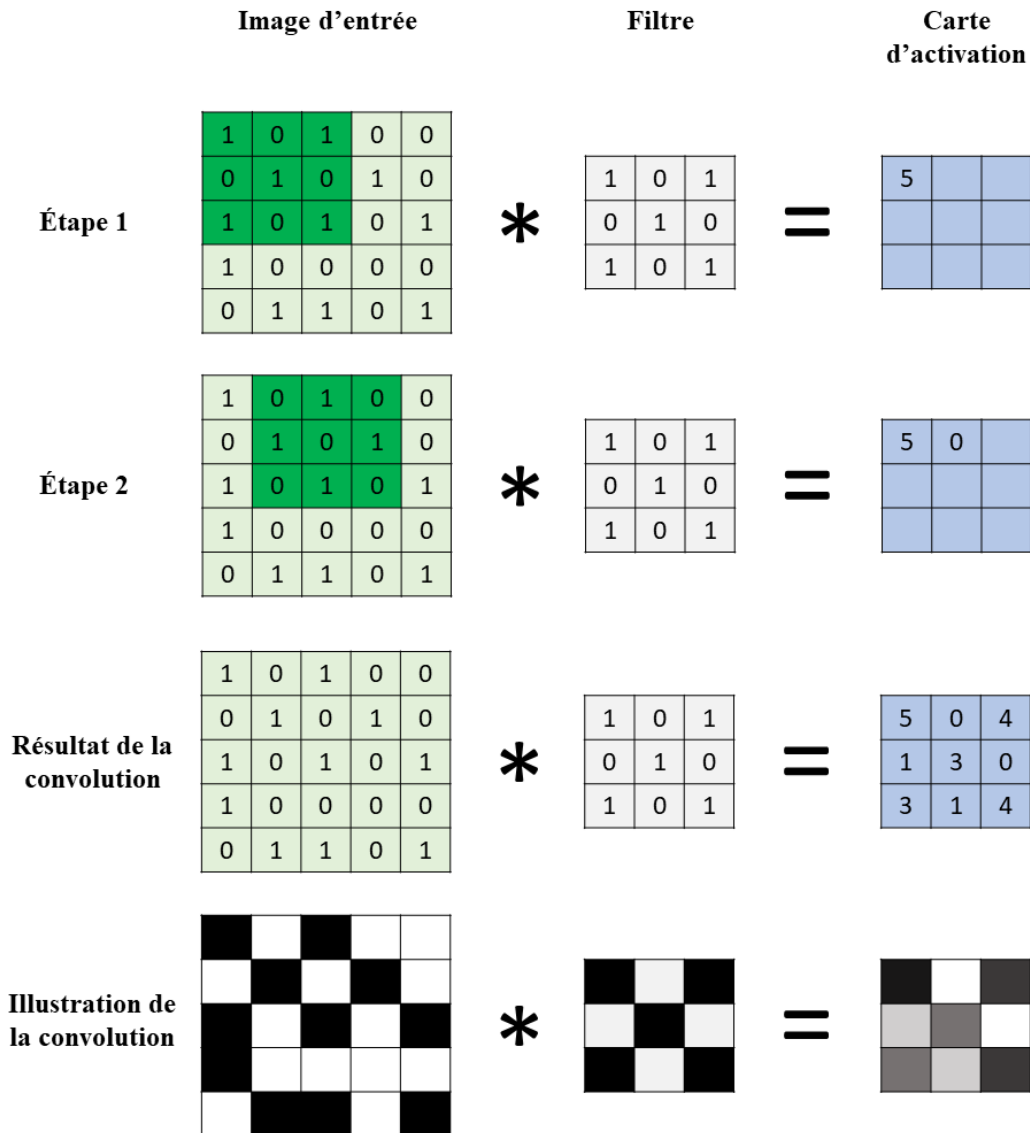


Figure 6. Exemple de convolution d'une image d'entrée. **Étape 1 :** Le filtre de taille 3×3 est superposé sur la région en haut à gauche de l'image d'entrée. Le produit scalaire entre le filtre et cette région d'image est calculé et les valeurs de la matrice résultante sont sommées. Le résultat obtenu est de 5. **Étape 2 :** Cette opération est réitérée après avoir glissé la fenêtre d'un cran vers la droite (pas de 1). Cette fois, le produit scalaire sera une matrice remplie de 0 ; aucune partie du filtre n'est détectée. La somme de cette matrice est donc 0. **Résultat de la convolution :** Ces opérations sont répétées pour chaque fenêtre de taille 3×3 de l'image. La matrice finale, présentée à droite, est la carte d'activation. **Illustration de la convolution :** Au lieu de lire les valeurs, il est possible d'imaginer l'intensité des pixels à la place. Ainsi le blanc représente la valeur 0, et le noir représente la valeur 1. Nous pouvons alors remarquer que le filtre détecte le degré de présence d'une croix. La croix présente tout en haut à gauche de l'image explique la valeur élevée (qui est de 5) tout en haut à gauche de la carte d'activation : le filtre détecte fortement une croix à cet endroit.

La dernière couche d'un modèle, la couche de sortie, est généralement un vecteur d'une taille variable. Pour un problème de classification binaire par exemple, la couche de sortie peut contenir une seule valeur correspondant à la probabilité d'une des deux classes. Pour un problème de classification sur n classes, la couche de sortie peut être un vecteur de taille n , où chaque valeur est la probabilité que l'image d'entrée appartienne à la classe correspondante. Cette couche fait partie des couches appelées *fully-connected* (entièrement connectée). Pour une couche *fully-connected*, chaque neurone sera une fonction d'une combinaison linéaire de tous les neurones de la couche précédente. Le choix de cette fonction, appelée fonction d'activation, dépend du problème à résoudre. Différentes fonctions d'activation existent, telles que la fonction Rectified Linear Unit (ReLU) (Unité Linéaire Rectifiée), définie telle que $ReLU(x) = \max(0, x)$, ou encore la fonction sigmoïde. Les couches *fully-connected* sont particulièrement utilisées à la fin de l'architecture des CNNs, après les couches de convolution et avant la couche de sortie.

Hormis ces couches, un autre type de couche est également très populaire : les couches de *pooling* (mise en commun). Les couches de *pooling* permettent de réduire considérablement la taille des images, sans perte majeure d'information. Elles prennent en entrée des cartes d'activation, et sont souvent placées entre deux couches de convolution. L'opération de *pooling* la plus utilisée est le *max pooling* (Figure 7). Dans ce cas, l'image est divisée en plusieurs parties (toutes de même dimension, généralement 2×2), et chacune de ces régions est remplacée par la valeur maximale qu'elle contient.

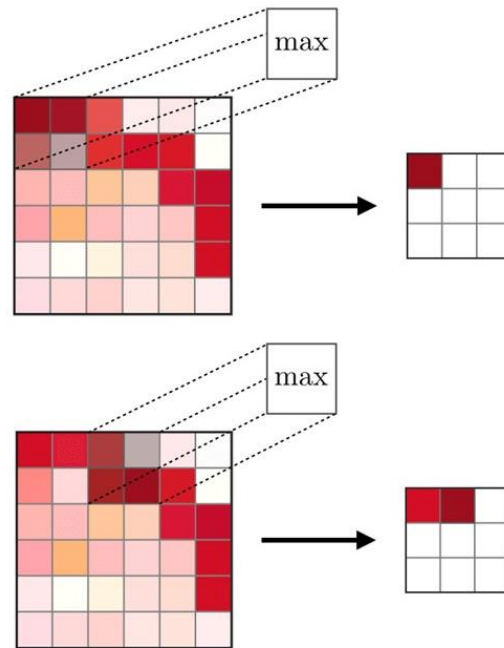


Figure 7. Exemple d'opération de max pooling. Ici, le pas est de 2 : chaque carré divisant l'image est disjoint. Dans ce cas, le nombre de pixels total est divisé par 4, adapté de <https://stanford.edu/~shervine/l/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels>

ii) Entraînement des modèles

Comme expliqué précédemment, les filtres des couches de convolution et les poids des neurones des couches fully-connected sont, au départ, aléatoires. Afin d'entraîner le modèle, le jeu de données est séparé en deux parties appelées jeu d'entraînement et jeu de validation. Le jeu d'entraînement contient généralement 70 % des données et sert, comme son nom l'indique, à entraîner le modèle. Lors de l'entraînement du modèle, une *fonction de perte* (loss function), également appelée fonction de coût, préalablement définie, permet de calculer l'erreur commise par le modèle sur des exemples annotés. Différentes fonctions de perte existent et permettent de calculer un score en utilisant la différence entre la réponse prédite et la réponse réelle, qui sera utilisé afin d'ajuster les paramètres du modèle. Pour cela, un algorithme, appelé *back-propagation* (rétropropagation du gradient), permet de calculer le gradient de tous les poids du modèle afin de quantifier la

contribution de chaque paramètre à la prédiction, relativement à la fonction de perte. Une fois ces gradients calculés, les poids du modèle sont mis à jour à l'aide d'un *optimiseur* (optimizer) qui minimise la fonction de perte et améliore les performances du modèle, à chaque itération. Le *taux d'apprentissage* (learning rate) est un hyperparamètre fondamental définissant à quel point les poids changeront en accord avec les gradients calculés, dans l'objectif d'atteindre un minimum de la fonction de perte. Plus le taux d'apprentissage sera élevé, plus le « pas » réalisé sur la fonction de perte sera important (voir Figure 8). Un taux d'apprentissage faible nécessitera généralement plus d'itérations pour converger, augmentant potentiellement le temps de calcul. Un taux d'apprentissage élevé risquera en revanche plus facilement de manquer un minimum local et donc de ne pas minimiser la fonction de perte. Il est courant que les poids du modèle ne soient pas mis à jour après chaque prédiction d'une image du jeu d'entraînement, mais plutôt d'un ensemble de ces images, appelé *batch* (lot). Le batch size (taille d'un batch) est un hyperparamètre à définir par l'utilisateur. Trouver la valeur optimale de ce paramètre est complexe et différentes valeurs sont habituellement testées.

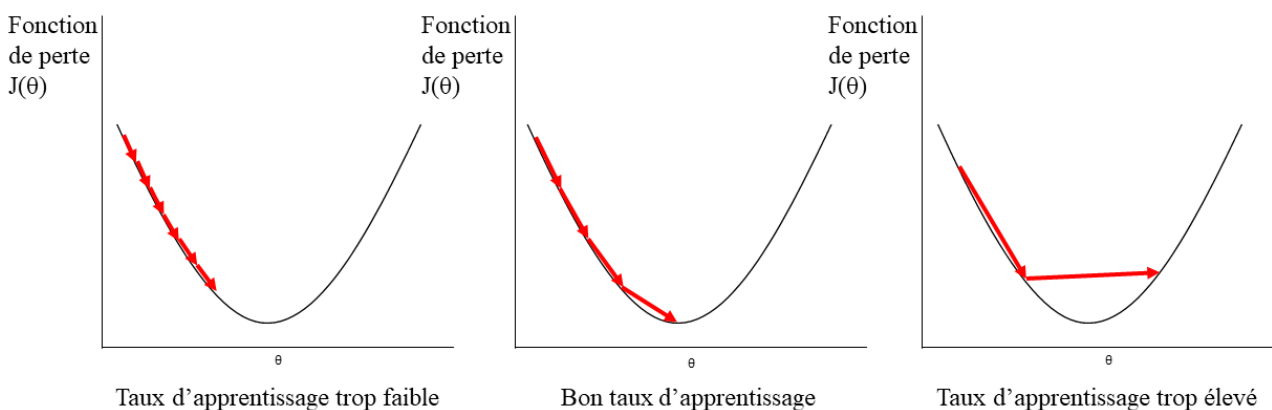


Figure 8. Représentation de différents taux d'apprentissage. Chaque flèche rouge correspond à une itération d'entraînement du modèle. À gauche, un taux d'apprentissage faible impliquera un nombre élevé d'epochs pour atteindre un minimum local. Au milieu, un bon taux d'apprentissage, permettant d'atteindre un minimum local de la fonction de perte en un nombre modéré d'epochs. Le taux d'apprentissage plus élevé sur la courbe de droite ne permet en revanche pas d'atteindre le minimum local.

Une fois que toutes les données d'entraînement ont été données au modèle et les poids mis à jour, cette opération est réitérée un certain nombre de fois. Ce nombre, appelé *nombre d'epochs*, est un hyperparamètre définissant le nombre de fois où le modèle sera entraîné sur le jeu d'entraînement. Généralement, le nombre d'epochs choisi est élevé, afin de permettre au modèle de capturer toute la complexité du jeu d'entraînement ainsi que la convergence du modèle. Après chaque epoch, les données de validation sont prédites par le modèle afin de calculer une métrique de performance du modèle. Généralement, la précision du modèle sur les données de validation est calculée, ainsi que la validation loss, qui correspond à la somme de la loss function calculée pour chaque image de validation. L'équivalent existe également sur la training loss (calculée cette fois sur le jeu d'entraînement). Les *courbes d'apprentissage* (learning curves) sont des représentations graphiques de la performance du modèle en fonction du nombre d'epochs. Représenter la validation et training loss selon les epochs, ou bien la précision de validation et d'entraînement selon les epochs permet de diagnostiquer les potentiels problèmes liés à l'étape d'entraînement (voir Figure 9), tels que l'underfitting et l'overfitting. De telles représentations sont appelées des courbes d'apprentissage.

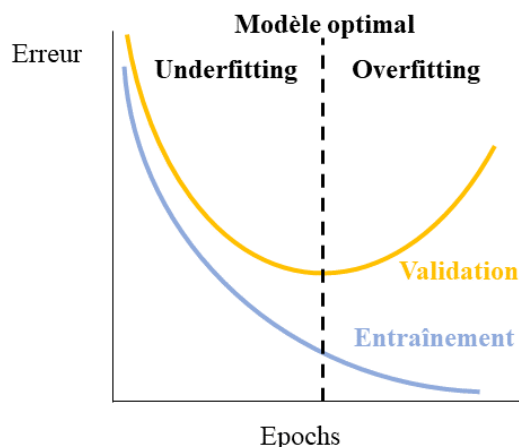


Figure 9. Exemple de courbe d'apprentissage. Les courbes, avant les pointillés, représentent un modèle victime d'underfitting : l'erreur du modèle diminue sur les données d'entraînement mais reste élevée pour l'ensemble de validation. Le modèle atteint ensuite son erreur la plus faible sur le jeu de validation, représenté par les pointillés. L'erreur d'entraînement continue ensuite de diminuer, tandis que l'erreur de validation augmente, caractérisant de l'overfitting : le modèle est alors trop dépendant des données d'entraînement et ne s'adapte plus à l'ensemble de validation.

c. Amélioration des performances d'un CNN

Bien que le big data soit omniprésent et ait transformé de nombreux secteurs, certains domaines ne permettent d'avoir qu'une quantité limitée de données, due à différents facteurs tels que des contraintes de confidentialité ou encore des limitations technologiques. Entraînés sur de grands jeux de données, les modèles de DL s'avèrent généralement très performants, mais sont bien plus limités quand la taille du jeu de données est amoindrie. En effet, les petits jeux de données sont sujets à différents biais et difficultés. Un premier problème, qui est valable également pour d'autres modèles que ceux de DL, est le manque de représentativité d'un tel jeu de données. Un petit jeu de données sera en effet beaucoup plus sensible aux données utilisées, omettant potentiellement certains motifs présents dans le monde réel. Deuxièmement, compte tenu du nombre important de paramètres des modèles de DL, il pourra être difficile d'optimiser efficacement le modèle et de capturer toutes les caractéristiques et tous les motifs présents dans les images. Enfin, les modèles de DL entraînés sur des petits jeux de données sont souvent sujets

à l'overfitting (surapprentissage). Ce phénomène survient quand un modèle est trop adapté aux données d'entraînement, capturant ainsi le bruit présent dans les données, et devient donc dans l'incapacité de généraliser les connaissances acquises. Imaginons un étudiant s'entraîner pour un partiel en refaisant sans cesse les mêmes annales. À force, il connaîtra par cœur le sujet et saura parfaitement y répondre, mais ne sera pas en mesure d'adapter ses connaissances face à un nouvel examen et de nouvelles données. L'overfitting consiste exactement en cela. Le modèle sera très performant sur les données d'entraînement, mais produira des résultats erronés sur de nouvelles données. L'underfitting (sous-apprentissage) est équivalent de l'overfitting quand le modèle, étant trop simpliste, n'apprend pas assez de l'ensemble d'entraînement. Pour toutes ces raisons, diverses techniques ont été développées, permettant de réduire ces biais et d'avoir des modèles performants, même entraînés sur des jeux de données de taille modeste.

Augmentation de données

Une des techniques les plus simples consiste à augmenter artificiellement, et de manière conséquente, la taille du jeu d'entraînement. En plus des images de cet ensemble, des transformations de ces dernières serviront pour l'entraînement. Les transformations peuvent être géométriques (rotation, symétrie, etc.) ou encore liées à la couleur (luminosité, contraste, etc.), mais il est fondamental que les transformations choisies soient pertinentes pour la tâche à résoudre. Cette technique, appelée augmentation de données, doit être exclusivement utilisée sur l'ensemble d'entraînement, et non sur l'ensemble de validation. À travers les images générées, cette technique permet, en plus d'améliorer la performance de modèles sur des petits jeux de données, de créer de la variabilité sur le jeu d'entraînement, rendant le modèle en question plus robuste aux variations.

Apprentissage par transfert

Au lieu d'entraîner un modèle en partant de zéro, il est possible d'utiliser des modèles disponibles en ligne et pré-entraînés sur des jeux de données déjà existants. Comme différents concepts en IA, cette idée peut se transposer chez les humains : une personne ayant joué du piano aura plus de facilité à apprendre la guitare qu'une personne n'ayant joué aucun instrument. La théorie, la pratique, les méthodes, la mémoire musculaire antérieurement acquises seront également utilisées, « transférées » pour apprendre un nouvel instrument. L'idée en DL est la même : bien qu'ayant été entraînée sur une tâche précise, la connaissance acquise par un modèle peut être transposée sur d'autres tâches relativement similaires. Ce concept, appelé apprentissage par transfert (transfer learning), peut être mis en place à travers différentes approches, selon la tâche à réaliser.

Il est par exemple possible de faire du *fine-tuning*, qui consiste à geler (fixer) certaines couches du modèle et le réentraîner. Par exemple, geler les premières couches du modèle (qui permettent de détecter des caractéristiques simples, telles que des lignes, des courbes...), puis le réentraîner sur les couches de plus haut niveau permet d'adapter le modèle à la nouvelle tâche de classification de manière plus efficace. La taille de la dernière couche peut être modifiée si le nombre de classes du modèle pré-entraîné et celui de la nouvelle tâche ne sont pas les mêmes. Il est également possible de rajouter des couches fully-connected.

Une autre application de l'apprentissage par transfert est l'extraction de caractéristiques. En partant d'un modèle pré-entraîné, l'objectif ici n'est pas de réentraîner le modèle, mais de faire de la prédiction sur une image afin d'en extraire les activations pour une couche donnée. Une fois toutes les extractions faites, chaque image peut être représentée par ses activations (représentant les

caractéristiques présentes dans l'image). Ces activations peuvent ensuite être utilisées afin d'entraîner des modèles plus classiques, tels qu'une régression logistique.

Early stopping

Lors de l'entraînement du modèle, il est possible de surveiller l'évolution des performances du modèle et la potentielle présence d'overfitting (voir Figure 9). L'early stopping est une technique permettant d'arrêter l'entraînement du modèle si, après un certain nombre d'epochs, une métrique prédéfinie a cessé de s'améliorer sur l'ensemble de validation. L'early stopping permet donc un compromis entre la complexité du modèle et sa capacité à généraliser sur de nouvelles données.

En plus de ces techniques, d'autres régularisations existent et permettent de réduire l'overfitting et d'améliorer la qualité du modèle en modifiant l'architecture d'un CNN.

Couche dropout

En effet, certaines couches peuvent être implémentées dans le modèle, telle qu'une couche dropout. Cette couche permet d'omettre aléatoirement certains neurones de la couche précédente. Ainsi, chaque neurone aura une probabilité p (prédéfinie par l'utilisateur) d'être omis à chaque batch. L'idée derrière cette technique est d'empêcher que certains neurones aient un poids trop important pour la classification et rendent la majorité des neurones obsolètes.

Régularisation L1 et L2

Comme pour la couche dropout, l'idée derrière les régularisations L1 et L2 est d'impacter le poids des neurones afin qu'aucun ne joue un rôle trop discriminant pour la classification. Les régularisations L1 (Eq. 1) et L2 (Eq. 2) consistent en une modification de la fonction de perte. Chaque méthode de régularisation présente

des avantages. La régularisation L1 favorise la parcimonie des paramètres (i.e. beaucoup de valeurs seront nulles, ce qui réduira la quantité de mémoire utilisée ainsi que le temps d'entraînement). Autrement dit, de nombreuses valeurs seront égales à 0, ce qui aura pour incidence de réduire la complexité et le nombre de caractéristiques sélectionnées par le modèle. La régularisation L2, quant à elle, permet également de réduire le temps de calcul.

$$Reg L1 = \text{Fonction de perte} + \lambda \sum_{j=1}^p |\omega_j| \quad (\text{Eq. 1})$$

$$Reg L2 = \text{Fonction de perte} + \lambda \sum_{j=1}^p \omega_j^2 \quad (\text{Eq. 2})$$

Batch Normalization

Une autre couche permettant de réduire l'overfitting et d'améliorer les performances du modèle est la couche de Batch Normalization. La normalisation consiste à soustraire à toutes les valeurs la moyenne puis de diviser ces nouvelles valeurs par l'écart-type. Ainsi, la moyenne des valeurs normalisées sera de 0 et l'écart-type de 1. C'est précisément l'objectif d'une couche de Batch Normalization, qui normalise les valeurs de sortie d'une couche donnée. Cette technique permet d'accélérer l'entraînement du modèle ainsi que de réduire le *covariate shift*. Ce phénomène survient quand la distribution entre le jeu d'entraînement et celle de validation sont différentes.

III. Problématique

1) Objectif

Nous avons vu que l'étude du comportement de dessin d'un point de vue phylogénétique et ontogénique a permis d'en apprendre beaucoup sur l'émergence du dessin ainsi que le développement cognitif chez l'enfant. Traditionnellement, les méthodes d'analyse du dessin consistent en premier lieu à définir les caractéristiques d'intérêt. Une fois ces caractéristiques définies vient l'étape d'extraction, qui peut être automatisée ou non selon leur complexité. Si par exemple le nombre d'yeux représentés sur un dessin est une caractéristique d'intérêt, il peut être difficile d'automatiser l'extraction de cette caractéristique. Il faudrait en effet pour cela avoir un modèle capable de détecter automatiquement la présence d'yeux, qui peut être une tâche complexe. Seulement, extraire cette caractéristique manuellement peut être complexe pour des dessins de jeunes enfants, où différents observateurs pourraient avoir une interprétation différente. Ce problème survient pour de nombreuses caractéristiques, plus abstraites, comme juger de l'agressivité ou la réalité d'une scène, ou plus généralement pour des dessins de primates non humains. D'ailleurs, que considérer comme caractéristiques d'intérêt pour ces productions ? Étant incapables de communiquer verbalement et comprendre les intentions d'un orang-outan, comment devrions-nous interpréter leurs productions ? Il est difficile de ne pas anthropomorphiser le processus du dessin en analysant le comportement de l'orang-outan ainsi que sa production.

Les biais cités précédemment peuvent être réduits à l'aide de mathématiques. En effet, l'utilisation d'indices interprétables et directement calculables sur des dessins permettrait d'extraire d'importantes informations contenues dans les dessins, sans

nécessité préalable d'analyser chaque dessin visuellement. Ces indices, bien que définis par des humains, restent plus objectifs que des variables liées à l'interprétation d'un dessin.

En plus de tels indices, les mathématiques peuvent être appliquées à travers des modèles de deep learning. En effet, de tels modèles permettent d'automatiser le processus d'extraction des caractéristiques, et ne nécessitent pas d'avoir préalablement défini des caractéristiques d'intérêt. Ainsi, les dessins analysés ne seront pas sujets à un jugement humain. Bien que ces modèles semblent idéaux pour analyser le dessin, ils sont en réalité très complexes et leur interprétation s'avère difficile. L'Article 1 (Chapitre 1) propose une revue de la littérature du deep learning appliqué au dessin ainsi que des perspectives d'applications.

Cette thèse s'inscrit dans un projet de recherche visant à rendre possible une étude du dessin plus objective à l'aide d'indices mathématiques. Le **Chapitre 2** définit et applique de tels indices au dessin. Le développement du dessin rend les formes plus complexes et plus détaillées. Seulement, cette complexité transparait-elle également dans l'aspect temporel de ce dernier ? Afin de répondre à cette question, l'Article 2 (Chapitre 2) propose, à travers l'application de fractales temporelles, d'étudier la complexité du processus de dessin, en formulant l'hypothèse que les dessins d'enfants seront moins complexes que ceux d'adultes (car non représentatifs, plus axés sur le geste que sur le dessin en lui-même). Les fractales temporelles sont également utilisées dans l'Article 3 (Chapitre 2), combinées à d'autres indices spatiaux et temporels, afin de comprendre comment ces derniers s'articulent entre eux, en formulant l'hypothèse que ces indices permettront de caractériser l'intentionnalité et la représentativité des dessins. Toujours chez l'humain, le **Chapitre 3** propose une approche basée cette fois sur le deep learning et les CNNs,

afin d'estimer la vitesse à laquelle le comportement de dessin se développe avec l'âge. L'étude menée dans l'Article 4 (Chapitre 3) se base sur des productions d'humains, enfants et adultes, afin de classer des dessins de différentes catégories d'âge, et de comparer différentes classes d'âge à l'aide des prédictions de CNNs. L'hypothèse formulée est que le modèle de deep learning sera d'une part capable de correctement classer des dessins d'enfants et adultes, et que d'autre part les prédictions permettront de quantifier le développement de ce comportement. Si le deep learning pouvait permettre d'en apprendre plus sur le comportement de dessin chez l'être humain, qu'en serait-il alors pour d'autres espèces? Pour tenter de répondre à cette question, le **Chapitre 4** s'intéresse à de telles productions, non figuratives, et difficiles, voire impossibles, à déchiffrer d'un œil humain. À l'aide de différents indices quantitatifs et qualitatifs, Pelé et al. (2021) ont démontré des différences saisonnières entre des dessins produits par une femelle orang-outan. Pouvons-nous retrouver ces différences, impossibles à deviner en observant simplement les dessins, à l'aide de deep learning? Pour répondre à cette question, l'Article 5 (Chapitre 4) utilise les mêmes dessins que Pelé et al. (2021), dans le but de classer les dessins selon leur saison de production, en laissant au modèle la tâche du choix et de l'extraction des caractéristiques, en émettant l'hypothèse qu'un tel modèle soit, dans une certaine mesure, capable de classer les dessins selon la saison de leur production. Afin d'en apprendre plus sur ce modèle, le style et le contenu des dessins sont également examinés plus en détail. L'attrait présenté par certains primates non humains pour le dessin questionne, mais il est difficile de savoir s'ils sont bien conscients de leur production et se pose alors la question de l'auteurisation. Dans la perspective de répondre à cette question, l'Article 6 (Chapitre 4) propose une approche réflexive et multidisciplinaire afin de trancher si

des animaux non humains ou des modèles d'intelligence artificielle devraient être considérés comme auteur ou non de leur production.

2) Article 1 : Et le deep learning appliqué au dessin ?

Deep learning for studying drawing behavior: A review

Publié dans *Frontiers in Psychology*

Synthèse de l'Article 1

Les réseaux neuronaux convolutifs sont des modèles adéquats et performants pour la classification d'images. Cette affirmation est en particulier vraie pour la classification de dessins. Cependant, il demeure difficile d'interpréter les facteurs considérés par de tels modèles. La première section de cet article de revue explore les différents travaux déjà existants sur l'application du deep learning au dessin en considérant le dessin en tant que processus et en tant que produit. Dans cette partie, de nouvelles applications des réseaux neuronaux convolutifs au comportement de dessin sont proposées à travers des approches ontogéniques ainsi que phylogénétiques. Enfin, la dernière partie est consacrée aux bases de données qui, combinées au deep learning, pourraient permettre d'en apprendre plus sur le comportement de dessin.

Deep learning for studying drawing behavior: A review

Benjamin Beltzung¹, Marie Pelé², Julien P. Renoult³, Cédric Sueur^{1,4}

¹ : Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

² : ANTHROPO LAB - ETHICS EA 7446, Université Catholique de Lille, F-59000 Lille, France

³ : CEFÉ, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

⁴ : Institut Universitaire de France, Paris, France

Corresponding author: Benjamin Beltzung

Email : benjamin.beltzung@iphc.cnrs.fr

Address: Institut Pluridisciplinaire Hubert Curien

Département Écologie, Physiologie et Éthologie

23, rue Becquerel

67087 Strasbourg

FRANCE

ORCID:

Benjamin Beltzung : 0000-0001-9437-7893

Marie Pelé : 0000-0003-2297-5522

Julien Renoult : 0000-0001-6690-0085

Cédric Sueur : 0000-0001-8206-2739

Acknowledgments

This project has received financial support from the CNRS through the MITI interdisciplinary programs and from the University of Strasbourg with an IDEX Exploratory Research program. We thank Editage for English editing service.

Statements and declarations

Funding

This project has received financial support from the CNRS through the MITI interdisciplinary programs and from the University of Strasbourg with an IDEX Exploratory Research program.

Competing interests

The authors have no competing interests to declare relevant to the contents of this article.

Abstract

In recent years, computer science has made major advances in understanding drawing behavior. Artificial intelligence, and more precisely deep learning, has displayed unprecedented performance in the automatic recognition and classification of large databases of sketches and drawings collected through touchpad devices. Although deep learning can perform these tasks with high accuracy, the way they are performed by the algorithms remains largely unexplored. Improving the interpretability of deep neural networks is a very active research area, with promising recent advances in understanding human cognition. Deep learning thus offers a powerful framework to study drawing behavior and the underlying cognitive processes, particularly in children and non-human animals, on whom knowledge is incomplete. In this literature review, we first explore the history of deep learning as applied to the study of drawing along with the main discoveries in this area, while proposing open challenges. Second, multiple ideas are discussed to understand the inherent structure of deep learning models. A non-exhaustive list of drawing datasets relevant to deep learning approaches is further provided. Finally, the potential benefits of coupling deep learning with comparative cultural analyses are discussed.

Keywords: deep learning, drawing behavior, sketch, artificial intelligence, cognition, primates.

Introduction

Drawing is a powerful communication medium that can convey concepts beyond words. Two different approaches are traditionally used to study drawing behavior (Pysal et al. 2021): the process approach (Adi-Japha et al. 1998; Freeman and Cox 1985) and the product approach (Brooks 2009; Xu et al. 2009). The process approach analyzes drawings through the behavioral characteristics linked to the drawing task and the individual who is drawing. For example, this perspective may require information on behavioral sequences (investigated through coordinates and the time spent drawing each point or behavioral sampling), which is more difficult to collect than the data needed for the product approach. Indeed, the latter analyzes the result of the drawing, based only on spatial and visual information, to infer the underlying behavior. Drawings, as final products, have been widely used to better understand the cognitive capacities of individuals, in particular to investigate the cognitive development of children (Barraza 1999; Cox 2005; Farokhi and Hashemi 2011; Malchiodi 1998). Studying visual features such as the color palette in drawings, the product approach has been pivotal in describing the diversity of personalities in children (Goldner and Scharf 2011), identifying mental disorders (Tharinger and Stark 1990) and post-traumatic symptoms (Backos and Samuelson 2017), and even revealing concealed emotions (Fury 1996). Both of these approaches - process and product - are covered in this review.

In toddlers, first drawings are in the form of scribbles, described as a motor activity not directed by the eyes, but by the mechanical functioning of the motor system arm-wrist-hand (Freeman 1993; Piaget and Inhelder 1967). At this age, scribblers appear to take little interest in their final products, whereby the process of drawing itself or improving the technique prevails over the will of representation (Golomb,

1992; Thomas and Silk, 1990). Figurative drawings, where what is drawn is representative for both the subject and external eyes, only appear at 3-4 years of age (Golomb, 1992; Freeman, 1993).

However, figuration and internal representativeness are not always similar. Since the end of the 19th century, researchers have developed a methodology to address the difficulties of studying drawings and scribbles (Farokhi and Hashemi 2011). These analyses are limited by the subjective judgment of the observer (Lark-Horovitz 1942), which is prone to several biases, especially with respect to semantic analyses. These issues are minor when computing low-level features such as color statistics, but are fundamental when trying to extract higher-level features; for example, one observer may see a house where another observer only sees a scribble, or both observers may fail to detect the drawer's intention to represent a house. The distinction between figuration and internal representativeness is essential, particularly when analyzing young children's drawings. Indeed, while previous theories proposed that the drawing among the youngest reflect motor activity only, recent studies have provided evidence for a symbolic function of drawing as early as 2 years old, suggesting that even young children can learn and become aware of the two visual aspects of drawing: the referent, which is the concept of what is drawn, and the signifier, which is the drawing object itself (Quaglia et al. 2015). However, a young child using drawings for symbolic representation may not intend to represent the formal aspects of reality through his or her first drawings, but rather seeks to express the world around him or her in a physiognomic way, using the line as means of expression (Quaglia et al. 2015). In other words, what is regarded as a scribble for an adult can be a symbolic representation for a young child. To understand the emergence and development of drawings, it is important to interpret such drawings.

To do so, asking very young children about their product is impossible, as they cannot communicate verbally. To address this problem, one could ask adults to interpret the drawings. However, by doing so, adults would typically fail to detect the intention of the drawer and the meaning of scribbles. Asking the child about his/her intention only partially solves this problem because for a given child, the answer has been shown to vary from one day to the next (Martinet et al. 2021). The answer is also dependent on the subject's verbal communication skills, which are naturally limited in toddlers, as in other great apes. This is not a problem for free-form drawings (i.e., no instruction), but becomes challenging for task-based drawings (i.e., instructions and constraints on the drawings) (Martinet et al. 2021). The same problem arises among great apes such as chimpanzees (*Pan troglodytes*), who are well known for their drawing behavior (Martinet and Pelé 2021). Indeed, captive chimpanzees spontaneously draw and paint if provided with appropriate materials (pen, paint, brushes, and paper) and can continue this behavior without being reinforced with food (Boysen et al. 1987; Tanaka et al. 2003).

To interpret the intention behind drawings, objective and mathematical analyses have been developed. Martinet et al. (2021) elaborated an innovative mathematical tool based on spatial fractal analysis, and Beltzung et al. (2022) used temporal fractal analysis for this purpose. The combination using a principal component analysis of simple metrics (number of lines, circles, colors, cover rate, etc.) can also provide interesting results regarding interindividual differences in human (Sueur et al., 2021) or orangutan drawings (Pelé et al., 2021).

The rise of deep learning

Over the last few decades, researchers have been investigating drawings using AI and computer vision (Eitz et al. 2012; Li et al. 2013). The latter encompasses sophisticated techniques and algorithms which can extract features in an image that are meaningful to human visual perception, such as facial features (e.g., eyes and nose). These techniques are widely used for detection (e.g., corner and edge detection (Li et al. 2015)), segmentation (e.g., K-mean, P-Tile), and recognition (convolutional neural network). Most analyses use computer vision to extract features which are then fed into a classifier.

It is important to note that traditional models and machine learning have been successfully used as approaches to study the drawing behavior. For example, by measuring the proportion of time the pen was in contact with the paper, Cohen et al. (2014) have shown a link between the Digital Clock-Drawing test and depression. Palsley et al. (2021) used machine learning methods, as Random Trees and Random Forest, to demonstrate how curvature and corners in drawings are linked to the age. These mathematical analyses and indices are objective contrary to former measures and are a good starting point for developing more objective studies using artificial intelligence (AI).

Currently, the most efficient and promising way to learn from images, including drawings, is deep learning (Fig. 1) (Ravindran, 2022), a sub-branch of computer vision and artificial intelligence, and more precisely neural networks, also used for speech recognition (Graves et al. 2013) and text classification (Liu et al. 2017). Deep learning allows us to go further by avoiding some anthropomorphic biases, such as the confirmation bias. For example, when analyzing drawings without deep learning, the features may be unconsciously selected accordingly to the beliefs of the human devising this process. By using almost raw data, deep learning thus reduces such biases.

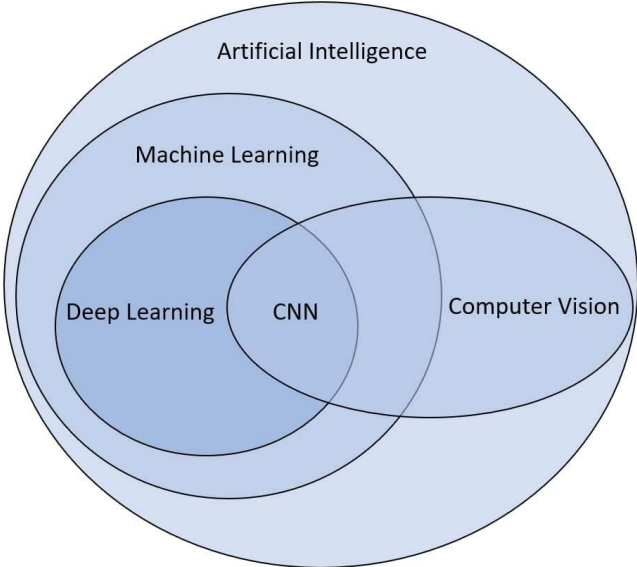


Figure 1. Euler diagram of artificial intelligence and neural networks in computer vision.

The first mathematical model defining the concept of artificial neurons dates back to McCulloch and Pitts (1943). Deep learning only surged in 2012, when a deep convolutional neural network (CNN) named *AlexNet* (Krizhevsky et al. 2012), outperformed other methods by a large margin in a popular competition of image classification, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC, (Deng 2009)). CNNs (Fig. 2) form a subcategory of artificial neural networks, specifically designed for processing images by learning filters (via convolutional layers) that

optimize performance in a predefined task (e.g., categorizing images or regressing images with a continuous variable). These filters allow capturing a hidden representation of images (Mukherjee and Rogers 2020). A glossary of technical terms is presented in Table 1.

Although deep learning is now a flagship approach to image analysis, most of these algorithms have been trained and designed for photos. Compared to photos, drawings and sketches are sparser and can be abstract. DL models thus need to be created to specifically process this type of data (Pysal et al. 2021; Yu et al. 2015; Zhang et al. 2016).

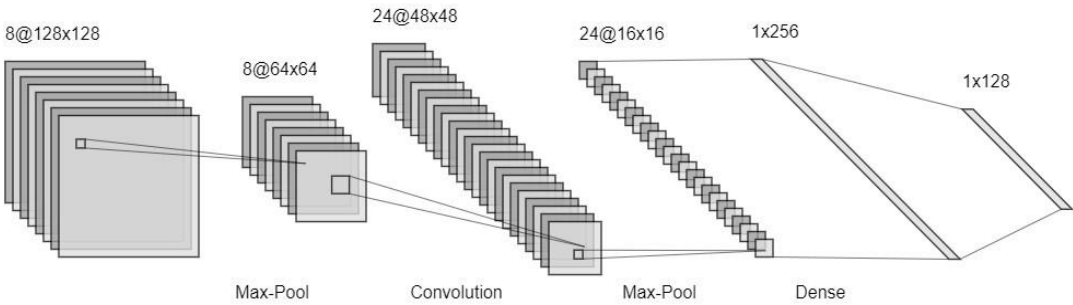


Figure 2. Example of a CNN architecture. The model takes an image as input; the image passes through layers to finally be classified between predefined classes.

These models can successfully classify drawings from several categories with high accuracy but allow limited interpretability. Indeed, deep learning models are often considered black boxes because of the number of parameters reaching tens of millions (Krizhevsky et al. 2012). Nevertheless, as in all scientific domains, interpretability and comprehension are key points when developing a model. What does a model outperforming human recognition ‘discover’ and ‘comprehend’ in the data that humans do not?

Table 1. Glossary of technical terms

Terms	Meaning	Definition
AI	Artificial Intelligence	All techniques allowing reproduction of intelligence.
ML	Machine Learning	Subset of AI techniques which learn from the training data.
DL	Deep Learning	Subset of ML techniques based on artificial neural networks. The analyzed features are learned by the model.
ANN	Artificial Neural Network	Model consisting of layers made up of units, also called neurons. An ANN can be shallow, or deep (DNN) when consisting of at least 2 hidden layers (i.e., layers between input and output).
CNN	Convolutional Neural Network	An ANN is specifically designed for images using convolutional layers.
RNN	Recurrent Neural Network	An ANN designed to process sequences, such as time series.
GNN	Graph Neural Network	An ANN where node relationships are studied.
GAN	Generative Adversarial Network	Unsupervised DL method capable of generating fake but realistic data.
VAE	Variational Autoencoder	An ANN belonging to the family of autoencoders, consisting of an encoder that compresses the data, and a decoder that reconstructs the data. Reconstruction in VAE is through a sampling of the hidden representation of the statistical model, rather than the hidden representation itself.
DA	Data augmentation	Techniques allowing an increase in the number of training data, by altering them through different transformations.
TL	Transfer Learning	Method consisting of reusing a model already trained on another task.

Is it possible to extract and decipher the discriminant features and are humans able to understand them? To improve the interpretability of these models and to answer these questions, multiple methods have been developed and are discussed later in this review. Nevertheless, interpretability and explainability remain important

challenges in deep learning (Gilpin et al. 2018) and are among the most active research topics in AI (Rudin 2019; Wu et al. 2018; Zhang et al. 2016). According to Gilpin et al. (2018), “the goal of *interpretability* is to describe the internals of a system in a way that is understandable to humans” and *explainability* (for deep networks) consists in giving an explanation to “the processing of data inside a network, or explaining the representation of data inside a network” (note that the definitions of these concepts are still debated, see for example (Tjoa and Guan 2021)). When studying drawings, the interpretability of AI is also fundamental to improving the knowledge of the ontogeny of drawing and the emergence of representativeness. Likewise, the AI processing of children and chimpanzees drawings can be compared to allow a better understanding of the evolutionary history of drawing. To achieve this goal, the assumptions on the underlying mechanisms of the drawing behavior can be formalized and implemented in a neural network model. With this objective, Philippsen and Nagai (2019) combined Bayesian inference and deep learning. They developed a neural network capable of completing partial drawings based on prior information. The goal of their study was to use this model to replicate children’s and chimpanzees’ drawing behavior to analyze the relative importance of different priors.

As previously mentioned, in children, the quality and representativeness of drawings improve with age (Martinet et al. 2021). In addition to age, other variables influence representation, such as sex (Picard and Boulhais 2011) and cultural background (Alland, 1983; Gernhardt et al. 2013). For example, Gernhardt et al. (2013) demonstrated that the number of facial details and facial expressions in drawings vary among children from different cultures. Deep learning is a promising tool for understanding cultural variations in drawing. To the best of our knowledge, no such

studies have been carried out yet. However, deep learning applied to drawings has recently been used to characterize mental disorders in individuals, such as Autism Spectrum Disorder (Anne et al. 2018), to predict the Draw-a-Person test (Widiyanto and Abuhasan, 2020), the Clock-Drawing test (Chen et al., 2020), and detect mild cognitive impairment (Ruengchaijatuporn et al., 2022).

Overall, deep learning in complement to other machine learning methods has the potential to greatly improve our knowledge of the ontogeny and evolutionary history of drawing behavior. This review presents and discusses the different applications of deep learning in drawing analysis and aims at giving the keys for readers who are interested by using deep learning to study drawing behavior and want to go further. The first section introduces different approaches to drawing analysis based on deep learning, which have already been applied or appear promising. These approaches are not discussed in relation to their performance (e.g., score of accuracy), but on the insights they can bring on the understanding of the drawing behavior. The second section reviews publicly available datasets that are well suited for studying drawings and sketches using AI and outlines the challenges. The review is concluded by discussing future research frameworks and perspectives in deep learning as applied to drawings.

Approaches in deep learning for drawing(s) analysis

This section is divided into two parts. The first part is focused on model-centric analyses, which refers to studies directly using the outputs of a model to make interpretation of the results. The second part focuses on analyses based on model-internals. The studies considered in this part use the weights of a model after being

trained (i.e., post hoc interpretation methods), such as heatmaps, to discover for example of the information is encoded in the model.

1. Model-centric analyses

As in classical drawing studies, deep learning approaches can be classified as focusing either on drawings as a product or on the process of generating drawings and sketches. While the first approach investigates only the spatial dimension of drawings, the second considers the temporal dimension.

1. Product approach

1.1 Prediction-based analyses

Machine learning models are often trained with the aim to predict labels for unlabeled data (i.e., that have never been seen by the model). In deep learning, this prediction task can be conducted at several levels, from labeling the image as a whole (classification) or predicting a label for every pixel of an image (segmentation).

1.1.1 Classification

The most popular application of deep learning is classification. Classification plays a major role in computer vision in tasks as varied as classifying Alzheimer's disease from magnetic resonance images (Wen et al. 2020), identifying fish species (Li et al. 2015), and recognizing malware images (Yue 2017). Classification is also a preliminary step in other tasks, such as segmentation. CNNs are mostly used for image classification in a supervised learning paradigm, where a model is trained to classify images into categories predefined by the user, by learning from a dataset of

labeled images (i.e., images for which the category is known). Once trained, the model is used to predict the categories of new, unlabeled images.

The first CNN developed for sketch classification was Sketch-a-Net (Yu et al. 2015), which achieved better performance than humans in object classification. It may be surprising that a model trained on data labeled by humans can outperform humans at classification. Indeed, CNNs learn a latent representation, that is, hidden features from the data, which is more complex than human representation. Sketch-a-Net performs better on sketches than neural networks trained on photos, highlighting the need for specific architectures for drawings (Yu et al. 2015). A CNN can thus outperform and replace classical methods used in sketch classification based on predefined classes. For example, in psychology, Vaitkevičius (2019), built a CNN capable of successfully classifying scribbles in 20 different classes as defined by the psychologist Rhoda Kellogg (e.g., “single dot,” “imperfect circle through single vertical line,” “spiral line”). Compared to other classifiers used in computer vision (e.g., support vector machine (SVM), random forest, k-nearest neighbors), CNN achieves the best results, matching the efficiency of neural networks in analyzing non-figurative drawings, demonstrating how deep learning can automatize complex and laborious tasks. Another example is Rakhmanov et al. (2020), who used a simple CNN architecture (two convolution layers and two fully connected layers) to classify drawings according to the Draw-a-Person test (Goodenough, 1926), a cognitive test in which the subject (a child, most often) draws a human figure, and a score is assigned to the drawing based on several criteria (e.g., the number of eyes, body proportions, presence of the mouth) to assess the child’s intellectual maturity. This test is used for several purposes, such as detecting behavioral disorders or measuring nonverbal intelligence. Several parameters are tested during the training of the

neural network mode, and data augmentation is applied to compensate for the low number of drawings, to significantly increase the accuracy. Data augmentation is a computer vision technique which is widely used in machine learning, which increases the size of the training data set by slightly modifying the original instances (that are images in this case, by applying rotation, horizontal flip, color contrasts for example) during the training phase. DA also reduces overfitting, a phenomenon that occurs when the model is too specialized for the training data and generalizes poorly on new data. Although the deep learning model was able to learn and produce relevant results, Rakhmanov et al. (2020) found that other methods of computer vision, such as the bag of visual words (BoVW) approach, outperformed CNN (62% accuracy for BoVW versus 52% for CNN). This example shows that a straightforward CNN design does not necessarily outperform state-of-the-art methods.

Thus, more complex CNN structures are required to learn the hidden representation of an object from sketches. To this end, Zhang et al. (2016) built *SketchNet*, a neural network capable of classifying sketches in object categories to discover the shared structures between real images and sketches belonging to the same category. The classification part of this model relies on associating a sketch image with a positive real image (from the same category) and a negative image (from another predicted category). The authors used an architecture consisting of three subnetworks optimized to extract features of the sketch images, positive images, and negative images. The features of the sketch and real images were eventually merged. *SketchNet* is based on prediction rankings. For a given sketch, the model computes the probability of the sketch belonging to each category, before returning the top five prediction categories (i.e., the five more likely predicted classes for this sketch) and the nearest real images.

1.1.2 Segmentation

Classification can be used as a preliminary step for other tasks, such as segmentation. Image segmentation partitions the pixels of an image into multiple regions and assigns a label to each pixel. This technique is widely used in various fields, such as medicine (Brzakovic and Neskovic, 1993) and video surveillance (Patro 2014). It can rely on classical computer approaches, but more recently, also on deep learning (Fig. 3).



Figure 3. Examples of segmentation results through CNNs from Chen et al. (2017). Reprinted with permission.

While classification helps improve segmentation, the opposite is also true. Sketches can be classified as a whole after segmentation and analysis of individual components, as in semantic sketch segmentation (SSS), which aims at labeling individual strokes. Semantic segmentation is notoriously difficult, however, because of complex perceptual laws, such as those proposed by Gestalt theory (Wertheimer, 1938). For example, the law of closure states that in an image with missing parts, the brain visually fills in the gaps. Interestingly, CNNs have been found to reproduce some perceptual laws. It appears that perceptual laws may or may not be present depending on the training set, and more generally the weights of the model (Kim et al., 2019, 2021; Jacob et al., 2021). For these reasons, it is a complex task to understand if and how neural network perception differs from that of humans, and these questions are still debated. Moreover, as with classification, it is necessary to

develop architectures and models of semantic segmentation for sketches, specifically because of the differences between sketches and photos.

One of the first CNN-based models of sketch segmentation was *SketchParse*, proposed by Sarvadevabhatla et al. (2017). *SketchParse* automatically parses regions of sketches and has proven to be effective, for example, in separating the head, body, and tail of a horse. However, *SketchParse* parses regions, not strokes, which limits the utility of segmentation in studying drawings as a process because regions most often are not consistent with strokes.

Graph neural networks (GNNs) can overcome this limitation. Starting from these neural networks, it is possible to cluster strokes into semantic object parts. Yang et al. (2021) proposed *SketchGNN*, a convolutional GNN which outperforms state-of-the-art models, such as SSS and stroke labeling. Their model also extracts features at three different scales: point-level, stroke-level, and sketch-level. *SketchGNN* can for example label each pixel of a sketch representing a face, to associate with the pixel a larger face component, such as the nose or the mouth, without taking into account the order of the strokes. Predicting object parts by strokes labeling could allow for comparing the structure of specific parts of an object depending on the culture of the drawers, for example to compare object proportions. Another interesting SSS model was proposed by Li et al. (2019). Their model is an hourglass-shaped network consisting of an encoder and a decoder. The 2D image passes through a network which predicts the segmentation map. The corresponding segmentation map is then transformed into a stroke-based representation, which is used to refine the segmentation map. Due to the lack of 2D annotated sketches, the network is trained on edge maps extracted from 3D models already segmented and labeled, which can thus be transformed into sketches. Moreover, as the model is trained on

3D models, several viewpoints are available, that may not be the ones frequently represented in drawings. As it would be questionable to analyze freehand sketches by using a network mainly trained on 3D model-transformed sketches, the authors evaluated their model on freehand sketch datasets (Eitz et al. 2012; Huang et al. 2014); their model outperformed previous ones. Comparing children's and 3D model-transformed sketches, for a given category, could improve our knowledge of their spatial representation.

1.2 Generation-based analyses

Deep learning used for image classification and segmentation is usually referred to as discriminative AI, where models are trained to convert high-dimensional inputs (e.g., images) into low-dimensional outputs (e.g., the names of depicted objects). In contrast, generative AI generates high-dimensional outputs (e.g., images) from low-dimensional inputs (e.g., semantic representations). Most people know generative AI through web-based applications that allow drawing one's portrait in Van Gogh's style or putting fake words in Obama's mouth on a video. However, beyond these applications, generative AI has become one of the most growing research areas in AI because of a very large array of applications (Bian and Xie 2021; Kell et al. 2020; Wu et al. 2016; Yang and Lerch 2020), which include a study of drawings.

1.2.1 Generation

Pallavi (2019) devised SuggestiveGAN, a generative adversarial network (GAN; Karras et al. 2019). A GAN is an unsupervised algorithm in which two neural networks compete. Fundamentally, one neural network (the discriminator) is a classifier to distinguish real images from fake images; the other neural network (the generator)

tries to generate the most ‘realistic’ fake images (according to the real dataset). SuggestiveGAN is able to reconstruct incomplete drawings (with missing strokes). The proposed model grasps the structure of the drawings at the expense of the details.

1.2.2 Style transfer

Style transfer involves applying the style of an image to another image, but not the content. In 2016, Gatys et al. (2016) proposed a CNN-based method of style transfer that quickly achieved high popularity owing to its impressive visual results. The method has been popularized by the famous Van Gogh painting, whose style has been widely transferred onto various kinds of portraits and landscape photos. The authors defined the style from the Gram matrix of activations, a measure of covariation between filters within a given layer (usually, all convolutional layers are used to define the style). The content is defined by the activation of the deepest convolutional layers. The stylized image is then obtained by searching a new image that simultaneously minimizes the distance between its content and that of the ‘content’ image, and the distance between its style and that of the ‘style’ image. One of the most famous examples of Gatys et al. (2016) model is the transfer of Van Gogh’s painting style to photographic portraits or landscapes.

Since the seminal work of Gatys et al. (2016), other CNN-based style transfer algorithms have been proposed and applied to various contexts (e.g., in user-assisted creation tools (Jing et al. 2020)). For drawings and sketches, it is necessary to design specific models of style transfer as in the classification of drawings which are sparser and have a higher level of abstraction compared to paintings. Chen et al. (2018) proposed *CartoonGAN*, a GAN-based style transfer algorithm developed for cartoon stylization. The model generates cartoon images based on real-world photos, which

can be useful for photo editing or for artists to gain time. More recently, Chen et al. (2018) proposed a framework capable of synthesizing face sketches while preserving details, such as skin texture and shading.

Hicsonmez et al. (2017) applied style transfer to drawings to learn the styles of different book illustrators. Their objective was to apply the style of drawing from an illustrator (the “style image”), to an image produced by a different illustrator (the “content image”). Their framework shows that this technique can be successfully applied to drawings. Dissociating the style and content of a drawing, and modeling how these two components vary separately would have numerous implications in drawing studies. For example, by using the style of children’s drawings, one may analyze the development of motor skills through the complexity of the strokes, by using only the style component of the drawing, while dissociating the motor constraints from the representational constraints. The style component can also be used to investigate the link between different types of curves used (broken curves and smooth curves) and internal representativeness (Adi-Japha et al., 1998). Moreover, studying the development of the style of the drawing system and the writing system, using style transfer, would help in understanding the differences and similarities between the two systems. Finally, using generative AI like the one developed by Chen et al. (2018), but instead generating realistic photos from drawings would shed light onto children’s representation of the world.

2. Process approach

2.1 Prediction-based analyses

2.1.1 Classification

In addition to the product approach, sketch recognition could allow a better understanding of the cognitive processes underlying the drawing. It is known that the development of drawing and writing shows kinematic differences according to age (Adi-Japha and Freeman 2001). Thus, classifying drawings and writing across ages could lead to discriminant low-level features, such as shapes, that could help in understanding the links as well as differences among techniques between these systems. Writing is not the only phenomenon correlated with drawings. Indeed, as shown by Panesi and Morra (2021), executive functions (e.g., shifting, inhibition) and language are linked to drawing behavior. Their work proposes several tasks to which different scores are assigned, such as the absence/presence of structures in human figure drawings, which can be further automated through deep learning. All these cognitive processes are directly linked to cortical activity, which is typically investigated using brain imaging (e.g., electroencephalography (EEG) and electromyography (EMG)). Applying deep learning to brain imaging can be achieved within a framework such as that proposed by Leandri et al. (2021) through recurrent neural networks (RNNs), which are specifically designed for temporal sequences.

He et al. (2017) developed a model able to use the temporal information of the strokes to perform sketch recognition as well as Sketch-based Image Retrieval (SBIR), which aims at finding real images visually similar to a given sketch. The proposed model is based on a CNN coupled with a R-LSTM (Residual Long Short-Term Memory) network. Multiple representations of the drawings are learnt by considering 60%,

80%, and 100% of the strokes of the drawings separately. The performance achieved by this model demonstrates how stroke ordering information can be used in deep learning and how it plays a role in classification. To go further, Xu et al. (2021) proposed to consider drawings as graphs thanks to GNNs (Graph Neural Network). A classical application of such architectures is node classification. A graph consists of edges and links, and GNNs analyze the relationships between the nodes. In sketches, these types of models take the relationships between the strokes into account. Their proposed model, called Multi-Graph Transformer, allows for capturing geometric and temporal information about the drawings, as well as understanding the relationship between strokes. These models could thus be useful to improve our knowledge on the links between object parts and object representations. These approaches could also help at understand which strokes are the most relevant for classification or comparing which parts of an object are drawn first depending on the culture or age for example.

2.1.2 Segmentation

The information contained in the stroke order and temporal sequences can provide very rich information, which may be hard to decipher just through image classification. Wu et al. (2018) designed a stroke-level sketch segmentation model, *Sketchsegnet*, that is based on a variational autoencoder (VAE) which learns the probability distribution from the data. In *Sketchsegnet*, widely used in image generation (Razavi et al. 2019; Zhu et al. 2020), the VAE consists of a bidirectional RNN (*BiRNN*) (Schuster and Paliwal, 1997) for the encoder and an autoregressive RNN (Inselberg and Dimsdale, 1990) for the decoder, thus accounting for the sequence order of strokes. For each sketch category, labels are predefined (e.g., ‘cream’ and ‘cone’ for an ice cream). Their model achieves an accuracy of 90% for stroke labeling.

Thus far, research on sketch and drawing segmentation using AI has been primarily methodological, with only rare applications to better understand the ontogeny and evolutionary history of drawing behavior. Nonetheless, segmentation could be of great interest in this kind of analysis. For example, segmentation can be used to analyze body proportions, which are indicative of the emotional state in children (e.g., disproportionally large hands can express aggression (Leo 2015), and the relative size of the head and trunk varies with age in children (Thomas and Tsalimi, 1988)). In addition, annotated sketch databases are not common, and annotating sketches will lead to bias, depending on the perception of the person doing the annotation. For this reason, SSS should be studied in depth through unsupervised stroke-level segmentation or by using temporal sequence algorithms (Gharghabi et al. 2019), which also consider the time spent not drawing. Applying SSS to scribbles could lead to semantic segmentation, not necessarily obvious to human perception. Moreover, SSS allows the analysis of specific regions, such as the head, at a certain level of detail despite the complexity. This could help in understanding the relative importance of different visual stimuli in shaping the representation space of children. Models using 3D sketches, similar to Li et al. (2019), can elucidate the emergence of 3D geometry in children, and more generally, the development of spatial ability in children, necessary for representativeness. Using deep learning to analyze low-level features such as the spatial distribution of strokes, their orientation and form, and how these vary with age could also be informative about the ontogeny of the drawing behavior in humans and other animals.

2.2 Generation-based analyses

As drawings are directly linked to the temporal sequences of the strokes, it is fundamental to consider the process when generating parts of drawings, to generate

meaningful strokes. Among the first to use generative AI were Ha and Eck (2017) who studied the behavior of drawing by developing a neural network capable of reproducing and mimicking human drawing through conditional and unconditional generation. To do so, they considered each drawing as a list of points, and each point as a vector of length 5 to characterize the position and state of the pen at a given time. The generative model used in this study was VAE. In Ha and Eck's (2017) model, both the encoder and decoder are recurrent neural networks, and hence, the name *Sketch-RNN*. When given an incomplete sketch, *Sketch-RNN* generates strokes to complete the sketch. As a result of the random nature of VAE, the model can predict different final results for the same initial sketch. The authors suggested that *Sketch-RNN* could be used, for example, to help students learn how to draw.

A model combining an RNN with Bayesian inference was developed by Philippsen and Nagai (2019) to unravel the sensory and cognitive mechanisms of drawing behavior. They relied on a 'predictive coding' scheme, according to which the brain constantly generates and updates internal, cognitive models of the world to predict the consequences of our actions in response to sensory inputs (Rao and Ballard, 1999). The authors investigated how varying the integration of sensory inputs with cognitive models influenced the ability of the RNN to learn to complete partial drawings. They found that a strong reliance on cognitive models is necessary to complete representational drawings, thereby highlighting the importance of internal models for efficient cognitive abilities such as abstraction and semantic categorization. Interestingly, the authors also stressed that drawings generated with a weak reliance on cognitive models differed from children's drawings but resembled chimpanzee's drawings. This result echoes previous suggestions that the inability of chimpanzees to complete representational drawings could be attributed to their poor predictive

cognitive skills, such as those involved in imagination (Saito et al. 2010; Watanabe 2013). This study also demonstrates the benefits of generative AI in understanding the development and evolution of drawing behavior. This predictive coding scheme can have other applications, such as understanding pathologies like metamorphopsia (e.g., straight lines that appear distorted) from the drawings of patients to unravel the neuronal mechanism that leads to these drawings.

II. Model internals-based methods

As we have seen, drawing behavior can be studied by designing and training deep neural networks models and directly interpret the output. However, these approaches do not take advantage of the internal knowledge learnt by the model. To address this issue, it is possible to develop techniques that use model internals, such as the weights and the neuronal activations of each layer separately.

Predictive models based on CNNs have been shown to outperform other models such as SVM and k-nearest neighbors, in most applications. However, as with any quantitative model, predictive power comes at the cost of interpretability, and a notorious limitation of CNNs is their low explanatory appeal (Rudin 2019). Regarding the ability of CNNs to help understand human behaviors, some researchers have suggested that AI is simply replacing a black box (the brain) with another. Other researchers have argued otherwise (Hasson et al. 2020). Ribeiro et al. (2016) developed a model to classify photographs of wolves and huskies. Based on accuracy alone, the model worked well. However, this model was in fact performing badly; all the pictures of wolves in the training set had snow in the background, and pictures of huskies did not. In learning the most discriminative features to separate images of wolves from those of huskies, the model thus focused on the presence or absence

of snow in the background and did not encode the features of these canines. This purposely bad-designed experiment highlights how the qualitative analysis of learned features can increase the model interpretability. CNNs have explicit architectural specifications; they are trained with user-defined learning rules; and one has direct access to the weights (strength of connections between neurons) and neuronal activation. Analyzing how varying these hyperparameters improves or deteriorates the fit between models and empirical data offers exciting venues of research, in exploring both the neuronal mechanisms of information processing and their behavioral expressions (e.g., Lindsay 2021; Richards et al. 2019). A remarkable example is the study by Philippsen and Nagai (2019) discussed previously, in which the authors varied the hyperparameters prior to analyzing the relative importance of sensory inputs and cognition in drawing behavior. More generally, when devising and training a model to discriminate between children and adult drawings, or between drawings of humans and other great apes, independent of model performance, one may be interested in knowing which features are responsible for AI discrimination. To do so, two possible approaches exist: local interpretation, allowing us to understand the features of a specific image (i.e., based on the data), and global interpretation, allowing us to understand class discrimination (i.e., based on the model).

1. Local interpretation

Local interpretation encompasses techniques aiming at understanding a specific prediction (i.e., for a given instance) for a given model. Applied to deep learning, such methods can help understanding which part of an image played a role for a given prediction task. This section will provide examples of such techniques applied to sketches.

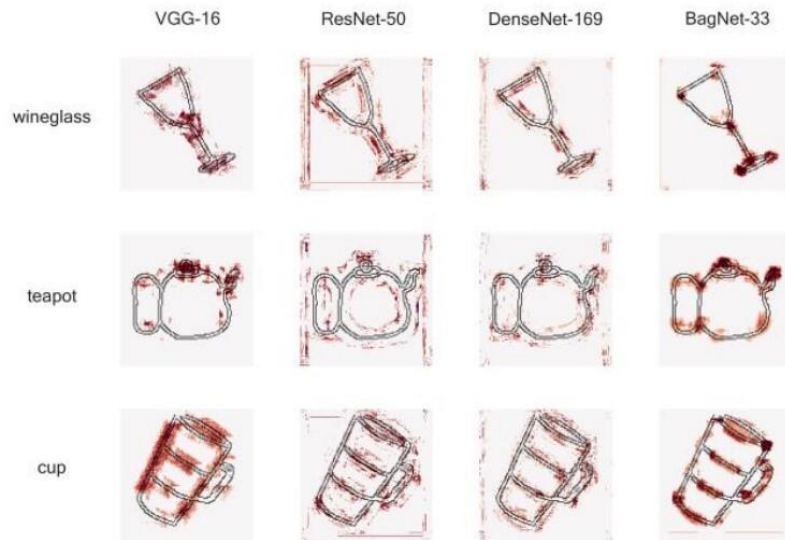


Figure 4. CAM of three objects by different models from Theodorus et al. (2020). The more the dots are clustered, the more of the corresponding area is considered in the model. Note that VGG-16 and BagNet-33 learned the representation of object parts. Reprinted with permission.

Bag-of-features (BoF) is a computer vision algorithm that aims to extract the occurrence count of features, and is more interpretable than CNNs. Brendel and Bethge (2019) developed *BagNet*, a neural network that combines the flexibility of CNNs and the interpretability of BoF. Although *BagNet* was originally created to analyze natural images, Theodorus et al. (2020) used this model to interpret sketch classification. They compared *BagNet* to non-interpretable CNNs such as VGG (Simonyan and Zisserman 2015), *ResNet* (He et al. 2016), and *DenseNet* (Huang et al. 2017), to determine whether the increased interpretability is due to the model itself or to the difference between natural images and sketches. *BagNet* was trained to classify sketches into 251 object categories. For each model (VGG, *ResNet*, *DenseNet*), the authors extracted and compared class activation maps (CAMs) (Zhou et al. 2016) for multiple images. For a given category, CAM indicates which region of an image influences the prediction of that category the most (Fig. 4). To go beyond the qualitative interpretation allowed by a simple description of CAMs, Theodorus et al. (2020) developed a quantitative metric of interpretability, the heatmap

interpretability (HI) score, which evaluates the quality of a CAM. A high HI score indicates a meaningful heatmap. Fig. 4 illustrates that the CAMs from *ResNet-50* and *DenseNet-109* have a low HI compared to *VGG-16* and *BagNet-33*, because highly activated pixels are largely scattered. Concurrently, a questionnaire was used to empirically evaluate the interpretability of the model. Each respondent was given one correctly predicted image per category with its corresponding heatmap and was asked to label object parts according to the heatmap. Comparing the CAMs for several categories, the authors concluded that their model did not use the same features as humans do for classifying object sketches. For example, the CAMs for the categories of ‘sword’ and ‘knife’ showed that the model only focused on the tip of these objects during classification, while humans also considered the handle and the shape of the blade.

Theodorus et al. (2020) provided an example of how interpretable deep learning models could be used for sketches. Although their model does not understand object representation as humans do (Baker et al. 2018; Jacob et al. 2021)), training models on different age classes separately and analyzing the heatmaps of several object categories can help formulate hypotheses about the development of drawing behavior in children. CNNs can be used in conjunction with eye tracking. By using the framework proposed by Theodorus et al. (2020), the dots from CAMs for a given object can be compared to those of humans when classifying an object. Eye-tracking and CNN can also be used for a phylogenetic approach to understand the visualization, understanding, and representation of objects of different apes, for a comparison with young children.

In addition to heatmaps, other techniques offer interpretations, such as perturbation-based models. An example of such a method is ZFNet (Zeiler and Fergus

2014), where parts of a given image are occluded and replaced by a gray square. Using this method, boxes can be occluded to understand which parts of the image are important for classification. However, it should be noted that the transparency of the prediction must be rigorously studied, as it may not be achievable through local interpretations (Ghassemi et al. 2021).

2. Global interpretation

To understand how information flows in the model, another possibility is to study the global interpretation, such as feature visualization. The first convolutional layer of a CNN extracts basic features, that is, edges and color blobs (Qin et al. 2018), which are easy to visualize and understand, while the deeper layers extract more complex shapes, which can describe parts of objects, entire objects, or complex patterns abstract to human perception. (Singer et al. 2020, 2021) showed that photographs and drawings are similarly represented in the early and intermediate layers for networks trained on photographs.

Feature visualization has been widely studied in computer vision (Olah et al. 2017; Yosinski et al. 2015; Zeiler and Fergus 2014), but few studies have been conducted on sketches and drawings. Young-Min (2019) studied the visual characteristics involved in comic book page classification. First, they designed a model to classify comic book pages between several comic artists. They then investigated visual features using a previously published method (Szegedy et al. 2015), with nine representative neurons for each layer. The results showed that, contrary to photograph classification, the features used by the CNN in classifying drawings of comics were not parts of objects, such as face features, but common artistic patterns (e.g., textures).

Applying these techniques to sketch classification, neural networks can discover new features for several classification problems, such as between very young children and chimpanzees, or even compare the drawing style between different cultures. The hierarchical order of the layer can also be meaningful in understanding the drawing behavior of children at several levels, ranging from a stroke to an object shape construction. For instance, when looking for interspecific and intraspecific differences in drawings, the first convolutional layer of a CNN can extract basic features, differentiating humans from other great apes. For a given CNN trained for classification, one can test whether the depth of the layer discriminant of the classifier is linked to the degree of behavioral divergence (that could be developmental, cultural, genetic, or phylogenetic). One would expect early layers to be discriminant enough to classify between species, and deeper layers would be needed for more complex classification, such as cultural or developmental divergences. As orangutans are more dexterous with their hands than chimpanzees and gorillas (Mackinnon, 1974), the first layer could separate humans and orangutans as well as the two other species.

Another way to interpret CNNs is by using the model parameters proposed by Chen et al. (2016), who developed InfoGAN, a generative model for interpretation that maximizes mutual information to discover latent features. This method has been evaluated using various datasets, such as the MNIST dataset (LeCun, 1998), a database of handwritten digits. In this case, the generator was able to discover latent features describing, for example, digit type, width, and rotation of the digits. InfoGAN has also been used on the CelebA face dataset (Liu et al. 2015), revealing encoded features like the azimuth, the presence of glasses, hairstyle, and emotion. From these results, we anticipate that InfoGAN would have a high appeal in studying

sketches, to explore the development of perception and representation in children by identifying features that are common and those that are discriminant between children and adult drawings.

Long et al. (2018) collected drawings from young children, older children, and adults to understand how object representation develops with age. They used a method called transfer learning, where a model trained on one task is reused for another task. Transfer learning saves computing resources and allows for high performance with a relatively small number of datasets because it exploits the fact that some properties learned by a model to solve one task are useful for many other related tasks. In the study by Long et al. (2018), sketches were encoded by a pre-trained CNN; features were extracted from layers across several depths; and representational dissimilarity matrices (RDMs) were calculated for each of the three-age classes and compared. Their study showed that the way older children represent objects is more similar to that of adults than young children. Moreover, this also raises the possibility of studying different levels of representation of drawings through different layers of CNNs. Thus, local and global interpretations are possible with CNNs.

Available datasets for drawing(s) analysis

As a result of the widespread availability of touch-screen devices, drawings and in particular, sketches, can now be more easily collected and analyzed. Moreover, scholars can also collect drawings online through crowdsourcing or online drawing games. However, these datasets have been rarely used in psychological or anthropological studies, possibly because of the lack of associated metadata on the

participants, such as their age, location, gender, culture, or drawing skill level. This metadata can be difficult to collect because they may require ethical approval.

Datasets can be organized into two families: amateur and expert datasets. In this review, amateur datasets mostly collate data on sketches and drawings without associated metadata on the person who did the drawing (in particular the drawing skills). Expert datasets gather drawings that have been extracted from books or comics. This kind of data can lead to other difficulties, such as copyrights. Moreover, the style difference between two experts may be significantly larger than that of between two amateurs, meaning that results obtained with one expert dataset may not be easily generalized to another expert dataset. We provide a non-exhaustive list of drawing datasets that are summarized in Table 2.

1. Amateur datasets

A major -and one of the first - sketch datasets is *QuickDraw* by Google. This dataset includes more than 50 million sketches belonging to over 345 object categories (Jongejan et al. 2016). *QuickDraw* is an online game where participants are asked to draw an object in 20 s, and a network is trained to recognize that object. For each sketch, the category is stored, as well as a Boolean indicating if the category was recognized by the game, the timestamp of the sketch, the country where the drawing was made, and the spatial and temporal data of the strokes. Despite some limitations (e.g., the lack of metadata such as the sex and age of the person who did the drawing), *QuickDraw* is a highly promising tool for investigating cultural differences in drawing-based object representations.

The second important amateur dataset is the TU-Berlin dataset, which provides more than 20,000 sketches of 250 categories of common objects drawn by 1,350 unique

participants (Eitz et al. 2012). TU-Berlin sketches were collected via Amazon Mechanical Turk (AMT), a crowdsourcing marketplace where requesters hire crowdworkers to perform particular tasks (in our case, drawing an object from a given category). Furthermore, each drawing is associated with a second category, for which other participants are asked to identify the drawn object. The temporal order of the strokes is available for each drawing; however, the personal data on the participants are not available.

A third dataset is the Sketchy database, which consists of 12,500 photographs of 75,471 sketches belonging to 125 object categories (Sangkloy et al. 2016). Each sketch is paired with a photograph, and each photograph is linked to a number from 1 to 5, characterizing the ease of sketching. Temporal data on strokes are available for each sketch. As most of the datasets are constructed by asking the participants to draw a particular object, there may be a large variability with respect to the drawn object and its features. For example, when asked to draw a dog, two participants may think about completely different breeds, which can be undesirable for the analyses. For this reason, datasets containing sketches representing photographs can lead to a decrease in variability, which can be an asset for this type of data.

II. Experts' datasets

Among expert datasets, Manga109 (Fujimoto et al. 2016) provides 109 Japanese comic books drawn by 94 professional creators with each book containing 194 pages on average. These books date from 1970 to 2010 and several genres are illustrated. Each page is annotated with rectangular areas characterizing the position of

metadata, such as frames, text, and character (face, body), through software developed for this study.

Hicsonmez et al. (2017) collected more than 6,500 pages from a total of 24 children’s book illustrators, with the goal of recognizing the authors using deep learning.

The list of datasets in this review is not exhaustive, only the main datasets are described. Other sketches datasets exist, such as COAD (Tirkaz et al., 2012), SPG (Li et al., 2018), SketchyScene (Zou et al., 2018).

Table 2. Summary of the presented datasets. All these datasets are at least partially available online

Dataset	# of images	# of categories of drawings	Drawing skill	Type of approach (product or process)	Metadata
<i>Quick, Draw!</i>	50 million sketches	345 (objects categories)	Amateur	Both	Country
<i>TU-Berlin dataset</i>	20,000 sketches	250 (object categories)	Amateur	Both	Human prediction available for each drawing
<i>Sketchy dataset</i>	75,471 sketches of 12,500 photographs	125 (object categories)	Amateur	Both	Drawings paired with a photograph
<i>Manga109 dataset</i>	109 Japanese comics with 194 pages each	94 (professional creators)	Expert	Product	Labeled rectangles around frames, faces....
<i>Hicsonmez et al. dataset (2017)</i>	6500 pages	24 (professional illustrators)	Expert	Product	

Future research framework and perspectives

This review provides an overview on how deep learning has been and could be used to increase our knowledge of drawing behavior. Understanding the ontogeny of drawing behavior has many fundamental applications including, diagnosis of pathologies and understanding perception. However, the classical methods used in psychology or comparative cognition, to analyze drawings, rely on verbalization by the author and the subjective interpretation of the experimenter, which limits the reproducibility of results; one way to overcome this is to use deep learning. Simple classification using deep learning can lead to high accuracy, but the interpretability and reliability of the input are not easy to assess, which is also true for supervised (classification, feature visualization) and unsupervised (InfoGAN) learning. Methods have been developed to interpret these results, such as heatmaps and similarity matrices, that are relevant to sketches. Another approach uses generative modeling (e.g., GANs) to generate drawings, to analyze the generative process, and eventually infer the underlying behavior. However, while drawing ontogeny is known to critically depend on various factors such as culture, age, and sex, the large datasets of drawings and sketches, currently used to train CNN and other AI algorithms, usually lack this kind of information. Thus, it is important to develop new datasets, methods, and criteria to advance our understanding of drawing behavior. A dataset with many ancillary variables could, for example, allow cultural analysis. By unraveling the extraordinary predictive capacity of models and through ongoing research to make these models more transparent, AI will undoubtedly significantly contribute to improving our understanding of the fundamental behavior of drawing, for humans and their relatives.

Declarations

This project received financial support from the CNRS through the MITI interdisciplinary programs and from the University of Strasbourg with an IDEX Exploratory Research program.

References

- Adi-Japha, E., and Freeman, N. H. (2001). Development of differentiation between writing and drawing systems. *Developmental Psychology* 37, 101-114. doi: <https://doi.org/10.1037/0012-1649.37.1.101>.
- Adi-Japha, E., Levin, I., and Solomon, S. (1998). Emergence of representation in drawing: The relation between kinematic and referential aspects. *Cognitive Development* 13, 25-51. doi: [http://dx.doi.org/10.1016/S0885-2014\(98\)90019-3](http://dx.doi.org/10.1016/S0885-2014(98)90019-3).
- Alland, A. (1983). *Playing With Form: Children Draw in Six Cultures*. Columbia University Press doi: 10.7312/alla92324.
- Anne, T., Philippsen, A., and Nagai, Y. Characterizing individual behaviors by using Recurrent Neural Networks. 15.
- Backos, A., and Samuelson, K. W. (2017). Projective drawings of mothers and children exposed to intimate partner violence: A mixed methods analysis. *Art Therapy* 34, 58-67. doi: <http://dx.doi.org/10.1080/07421656.2017.1312150>.
- Baker, N., Lu, H., Erlikhman, G., and Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology* 14, e1006613. doi : <https://doi.org/10.1371/journal.pcbi.1006613>.
- Barraza, L. (1999). Children's drawings about the environment. *Environmental education research* 5, 49-66. doi: <http://dx.doi.org/10.1080/1350462990050103>.
- Beltzung, B., Martinet, L., MacIntosh, A. J. J., Meyer, X., Hosselet, J., Pelé, M., et al. (2021). To draw or not to draw: understanding the temporal organization of drawing behaviour using fractal analyses. 2021.08.29.458053. doi: <https://doi.org/10.1101/2021.08.29.458053>.
- Bian, Y., and Xie, X.-Q. (2021). Generative chemistry: drug discovery with deep learning generative models. *J Mol Model* 27, 71. doi: <https://doi.org/10.1007/s00894-021-04674-8>.
- Boysen, S. T., Berntson, G. G., and Prentice, J. (1987). Simian scribbles: A reappraisal of drawing in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology* 101, 82. doi: <http://dx.doi.org/10.1037/0735-7036.101.1.82>.
- Brendel, W., and Bethge, M. (2019). Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet. *arXiv:1904.00760 [cs, stat]*. Available at: <http://arxiv.org/abs/1904.00760> [Accessed November 10, 2021].
- Brooks, M. (2009). Drawing, visualisation and young children's exploration of "big ideas." *International Journal of Science Education* 31, 319-341. doi: <http://dx.doi.org/10.1080/09500690802595771>.
- Brzakovic, D., and Neskovic, M. (1993). Mammogram screening using multiresolution-based image segmentation. *Int. J. Patt. Recogn. Artif. Intell.* 07, 1437-1460. doi: <https://doi.org/10.1142/S0218001493000704>.
- Chen, C., Tan, X., and Wong, K.-Y. K. (2018a). Face Sketch Synthesis with Style Transfer Using Pyramid Column Feature. in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 485-493. doi: <https://doi.org/10.1109/WACV.2018.00059>.
- Chen, S., Stromer, D., Alabdallah, H. A., Schwab, S., Weih, M., and Maier, A. (2020). Automatic dementia screening and scoring by applying deep learning on clock-drawing tests. *Sci Rep* 10, 20854. doi: 10.1038/s41598-020-74710-9.

- Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.). Available at: <https://proceedings.neurips.cc/paper/2016/hash/7c9d0b1f96aebd7b5eca8c3eadaa19ebb-Abstract.html> [Accessed November 10, 2021].
- Chen, Y., Lai, Y.-K., and Liu, Y.-J. (2018b). CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. in, 9465-9474. doi: <https://doi.org/10.1109/CVPR.2018.00986>.
- Cohen, J., Penney, D. L., Davis, R., Libon, D. J., Swenson, R. A., Ajilore, O., et al. (2014). Digital clock drawing: Differentiating 'thinking' versus 'doing' in younger and older adults with depression. *J Int Neuropsychol Soc* 20, 920-928. doi: [10.1017/S1355617714000757](https://doi.org/10.1017/S1355617714000757).
- Cox, S. (2005). Intention and meaning in young children's drawing. *International Journal of Art & Design Education* 24, 115-125. doi: <http://dx.doi.org/10.1111/j.1476-8070.2005.00432.x>.
- Deng, J. (2009). W., Dong, R. Socher, L. in J., Li,, K., Li,, and, L., Fei, Fei., *Imagenet: A large-scale hierarchical image database.*, In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages, 248-255. doi: <http://dx.doi.org/10.1109/CVPR.2009.5206848>.
- Eitz, M., Hays, J., and Alexa, M. (2012). How do humans sketch objects? *ACM Transactions on graphics (TOG)* 31, 1-10. doi: <http://dx.doi.org/10.1145/2185520.2185540>.
- Farokhi, M., and Hashemi, M. (2011). The analysis of children's drawings: social, emotional, physical, and psychological aspects. *Procedia-Social and Behavioral Sciences* 30, 2219-2224. doi: <http://dx.doi.org/10.1016/j.sbspro.2011.10.433>.
- Freeman, N. H. (1993). Drawing: Public instruments of representation.
- Freeman, N. H., and Cox, M. V. (1985). *Visual order: The nature and development of pictorial representation*. Cambridge University Press.
- Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., and Aizawa, K. (2016). Manga109 dataset and creation of metadata. in *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding* (Cancun Mexico: ACM), 1-5. doi: [10.1145/3011549.3011551](https://doi.org/10.1145/3011549.3011551).
- Fury, G. S. (1996). *The relation between infant attachment history and representations of relationships in school-aged family drawings*. University of Minnesota.
- Gatys, L. A., Ecker, A. S., and Bethge, M. (2016). Image Style Transfer Using Convolutional Neural Networks. in, 2414-2423. doi: <http://dx.doi.org/10.1109/CVPR.2016.265>.
- Gernhardt, A., Rübeling, H., and Keller, H. (2013). "This Is My Family": Differences in Children's Family Drawings Across Cultures. *Journal of Cross-Cultural Psychology* 44, 1166-1183. doi: <https://doi.org/10.1177/0022022113478658>.
- Gharghabi, S., Yeh, C.-C. M., Ding, Y., Ding, W., Hibbing, P., LaMunion, S., et al. (2019). Domain agnostic online semantic segmentation for multi-dimensional time series. *Data Min Knowl Disc* 33, 96-130. doi: <https://doi.org/10.1007/s10618-018-0589-3>.
- Ghassemi, M., Oakden-Rayner, L., and Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, e745-e750. doi: [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9).

- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (IEEE), 80-89. doi: <http://dx.doi.org/10.1109/DSAA.2018.00018>.
- Goldner, L., and Scharf, M. (2011). Children's family drawings: A study of attachment, personality, and adjustment. *Art Therapy* 28, 11-18. doi: <http://dx.doi.org/10.1080/07421656.2011.557350>.
- Golomb, C. (1992). The creation of a pictorial world.
- Goodenough, F. L. (1926). *Measurement of intelligence by drawings*. Oxford, England: World Book Co.
- Graves, A., Mohamed, A., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. in *2013 IEEE international conference on acoustics, speech and signal processing (IEEE)*, 6645-6649. doi: <http://dx.doi.org/10.1109/ICASSP.2013.6638947>.
- Ha, D., and Eck, D. (2017). A Neural Representation of Sketch Drawings. *arXiv:1704.03477 [cs, stat]*. Available at: <http://arxiv.org/abs/1704.03477> [Accessed November 10, 2021].
- Hasson, U., Nastase, S. A., and Goldstein, A. (2020). Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron* 105, 416-434. doi: <https://doi.org/10.1016/j.neuron.2019.12.002>.
- He, J.-Y., Wu, X., Jiang, Y.-G., Zhao, B., and Peng, Q. (2017). Sketch recognition with deep visual-sequential fusion model. in *Proceedings of the 25th ACM international conference on Multimedia*, 448-456.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. in, 770-778. doi: <http://dx.doi.org/10.1109/CVPR.2016.90>.
- Hicsonmez, S., Samet, N., Sener, F., and Duygulu, P. (2017). DRAW: Deep Networks for Recognizing Styles of Artists Who Illustrate Children's Books. in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval ICMR '17*. (New York, NY, USA: Association for Computing Machinery), 338-346. doi: <https://doi.org/10.1145/3078971.3078982>.
- Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. in, 4700-4708. doi: <http://dx.doi.org/10.1109/CVPR.2017.243>.
- Huang, Z., Fu, H., and Lau, R. W. H. (2014). Data-driven segmentation and labeling of freehand sketches. *ACM Trans. Graph.* 33, 175:1-175:10. doi: <https://doi.org/10.1145/2661229.2661280>.
- Inselberg, A., and Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. in *Proceedings of the First IEEE Conference on Visualization: Visualization 90*, 361-378. doi: [10.1109/VISUAL.1990.146402](https://doi.org/10.1109/VISUAL.1990.146402).
- Jacob, G., Pramod, R. T., Katti, H., and Arun, S. P. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. *Nat Commun* 12, 1872. doi: <https://doi.org/10.1038/s41467-021-22078-3>.
- Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., and Song, M. (2020). Neural Style Transfer: A Review. *IEEE Transactions on Visualization and Computer Graphics* 26, 3365-3385. doi: <https://doi.org/10.1109/TVCG.2019.2921336>.
- Jongejan, J., Rowley, H., Kawashima, T., Kim, J., and Fox-Gieg, N. (2016). The quick, draw!-ai experiment. *Mount View, CA*, accessed Feb 17, 4.

- Karras, T., Laine, S., and Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. in, 4401-4410. doi: <http://dx.doi.org/10.1109/CVPR.2019.00453>.
- Kell, D. B., Samanta, S., and Swainston, N. (2020). Deep learning and generative methods in cheminformatics and chemical biology: navigating small molecule space intelligently. *Biochemical Journal* 477, 4559-4580. doi: [10.1042/BCJ20200781](https://doi.org/10.1042/BCJ20200781).
- Kim, B., Reif, E., Wattenberg, M., and Bengio, S. (2019). *Do Neural Networks Show Gestalt Phenomena? An Exploration of the Law of Closure*.
- Kim, B., Reif, E., Wattenberg, M., Bengio, S., and Mozer, M. C. (2021). Neural Networks Trained on Natural Scenes Exhibit Gestalt Closure. *Comput Brain Behav* 4, 251-263. doi: [10.1007/s42113-021-00100-7](https://doi.org/10.1007/s42113-021-00100-7).
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25, 1097-1105. doi: <http://dx.doi.org/10.1145/3065386>.
- Lark-Horovitz, B. (1942). Comparison of subjective and objective judgments of children's drawings. *The Journal of Experimental Education* 10, 153-163. doi: <http://dx.doi.org/10.1080/00220973.1942.11010247>.
- Leandri, G., Schenone, A., and Leandri, M. (2021). Detection of movement related cortical potentials in freehand drawing on digital tablet. *Journal of Neuroscience Methods* 360, 109231. doi: <https://doi.org/10.1016/j.jneumeth.2021.109231>.
- LeCun, Y. (1998). THE MNIST DATABASE of handwritten digits. <http://yann.lecun.com/exdb/mnist/>.
- Leo, J. H. D. (2015). *Children's Drawings As Diagnostic Aids*. Routledge.
- Li, K., Pang, K., Song, J., Song, Y.-Z., Xiang, T., Hospedales, T. M., et al. (2018). Universal sketch perceptual grouping. in *Proceedings of the european conference on computer vision (ECCV)*, 582-597.
- Li, L., Fu, H., and Tai, C.-L. (2019). Fast Sketch Segmentation and Labeling With Deep Learning. *IEEE Computer Graphics and Applications* 39, 38-51. doi: <https://doi.org/10.1109/MCG.2018.2884192>.
- Li, X., Shang, M., Qin, H., and Chen, L. (2015a). Fast accurate fish detection and recognition of underwater images with Fast R-CNN. in *OCEANS 2015 - MTS/IEEE Washington*, 1-5. doi: <https://doi.org/10.23919/OCEANS.2015.7404464>.
- Li, Y., Song, Y.-Z., and Gong, S. (2013). Sketch Recognition by Ensemble Matching of Structured Features. in *BMVC*, 2. doi: <http://dx.doi.org/10.5244/C.27.35>.
- Li, Y., Wang, S., Tian, Q., and Ding, X. (2015b). A survey of recent advances in visual feature detection. *Neurocomputing* 149, 736-751. doi: <http://dx.doi.org/10.1016/j.neucom.2014.08.003>.
- Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience* 33, 2017-2031. doi: https://doi.org/10.1162/jocn_a_01544.
- Liu, J., Chang, W.-C., Wu, Y., and Yang, Y. (2017). Deep learning for extreme multi-label text classification. in *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, 115-124. doi: <http://dx.doi.org/10.1145/3077136.3080834>.
- Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep Learning Face Attributes in the Wild. in, 3730-3738. doi: <http://dx.doi.org/10.1109/ICCV.2015.425>.
- Long, B., Fan, J. E., and Frank, M. C. (2018). Drawings as a window into developmental changes in object representations. *Proceedings of the 40th*

- Annual Conference of the Cognitive Science Society*. Available at: <https://par.nsf.gov/biblio/10128363-drawings-window-developmental-changes-object-representations> [Accessed November 10, 2021].
- Mackinnon, J. (1974). The behaviour and ecology of wild orang-utans (*Pongo pygmaeus*). *Animal Behaviour* 22, 3-74. doi: [https://doi.org/10.1016/S0003-3472\(74\)80054-0](https://doi.org/10.1016/S0003-3472(74)80054-0).
- Malchiodi, C. A. (1998). *Understanding children's drawings*. Guilford Press.
- Martinet, L., and Pelé, M. (2020). Drawing in nonhuman primates: What we know and what remains to be investigated. *Journal of Comparative Psychology*. doi: <http://dx.doi.org/10.1037/com0000251>.
- Martinet, L., Sueur, C., Hirata, S., Hosselet, J., Matsuzawa, T., and Pelé, M. (2021). New indices to characterize drawing behavior in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Scientific reports* 11, 1-14. doi: <http://dx.doi.org/10.1038/s41598-021-83043-0>.
- McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5, 115-133. doi: <http://dx.doi.org/10.1007/BF02478259>.
- Mukherjee, K., and Rogers, T. T. (2020). Finding meaning in simple sketches: How do humans and deep networks compare? *Journal of Vision* 20, 1026-1026. doi: <http://dx.doi.org/10.1167/jov.20.11.1026>.
- Olah, C., Mordvintsev, A., and Schubert, L. (2017). Feature Visualization. *Distill* 2, e7. doi : <https://doi.org/10.23915/distill.00007>.
- Pallavi, S. (2019). Suggestive GAN for supporting Dysgraphic drawing skills. *IJ-AI* 8, 132. doi: <https://doi.org/10.11591/ijai.v8.i2.pp132-143>.
- Panesi, S., and Morra, S. (2021). Executive Function, Language, and the Toddler's Discovery of Representational Drawing. *Frontiers in Psychology* 12, 1926. doi: <https://doi.org/10.3389/fpsyg.2021.659569>.
- Patro, B. N. (2014). Design and implementation of novel image segmentation and BLOB detection algorithm for real-time video surveillance using DaVinci processor. in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1909-1915. doi: <https://doi.org/10.1109/ICACCI.2014.6968360>.
- Pelé, M., Thomas, G., Liénard, A., Eguchi, N., Shimada, M., and Sueur, C. (2021). I wanna draw like you: Inter- and intra-individual differences in orang-utan drawings. *Animals* 11, 3202. doi: <https://doi.org/10.3390/ani11113202>.
- Philippsen, A., and Nagai, Y. (2019). A predictive coding model of representational drawing in human children and chimpanzees. in *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 171-176. doi: <https://doi.org/10.1109/DEVLRN.2019.8850701>.
- Piaget, J., and Inhelder, B. (1967). The coordination of perspectives. *The child's conception of space* 8, 209-246.
- Picard, D., and Boulhais, M. (2011). Sex differences in expressive drawing. *Personality and Individual Differences* 51, 850-855. doi: <https://doi.org/10.1016/j.paid.2011.07.017>.
- Polsley, S., Powell, L., Kim, H.-H., Thomas, X., Liew, J., and Hammond, T. (2021). Detecting Children's Fine Motor Skill Development using Machine Learning. *International Journal of Artificial Intelligence in Education* 32. doi: [10.1007/s40593-021-00279-7](https://doi.org/10.1007/s40593-021-00279-7).
- Pysal, D., Abdulkadir, S. J., Shukri, S. R. M., and Alhussian, H. (2021). Classification of children's drawing strategies on touch-screen of seriation objects using a

- novel deep learning hybrid model. *Alexandria Engineering Journal* 60, 115-129. doi: <http://dx.doi.org/10.1016/j.aej.2020.06.019>.
- Qin, Z., Yu, F., Liu, C., and Chen, X. (2018). How convolutional neural network see the world - A survey of convolutional neural network visualization methods. *arXiv:1804.11191 [cs]*. doi : <http://dx.doi.org/10.3934/mfc.2018008>.
- Quaglia, R., Longobardi, C., and Iotti, N. O. (2015). Reconsidering the scribbling stage of drawing: a new perspective on toddlers' representational processes. *Frontiers in psychology* 6, 1227.
- Rakhmanov, O., Agwu, N. N., and Adeshina, S. (2020). Experimentation on Hand Drawn Sketches by Children to Classify Draw-a-Person Test Images in Psychology. in *The Thirty-Third International Flairs Conference* Available at: <https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS20/paper/view/18457> [Accessed November 10, 2021].
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2, 79-87. doi: <https://doi.org/10.1038/4580>.
- Ravindran, S. (2022). Five ways deep learning has transformed image analysis. *Nature* 609, 864-866. doi: [10.1038/d41586-022-02964-6](https://doi.org/10.1038/d41586-022-02964-6).
- Razavi, A., van den Oord, A., and Vinyals, O. (2019). Generating Diverse High-Fidelity Images with VQ-VAE-2. in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.). Available at: <https://proceedings.neurips.cc/paper/2019/hash/5f8e2fa1718d1bbcadf1cd9c7a54fb8c-Abstract.html> [Accessed November 10, 2021].
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv:1602.04938 [cs, stat]*. doi: <http://dx.doi.org/10.18653/v1/N16-3020>.
- Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nature neuroscience* 22, 1761-1770. doi: <https://doi.org/10.1038/s41593-019-0520-2>.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 1, 206-215. doi: <https://doi.org/10.1038/s42256-019-0048-x>.
- Ruengchaijatuporn, N., Chatnuntawech, I., Teerapittayanon, S., Sriswasdi, S., Itthipuripat, S., Hemrungronj, S., et al. (2022). An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks. *Alzheimer's Research & Therapy* 14, 111. doi: [10.1186/s13195-022-01043-2](https://doi.org/10.1186/s13195-022-01043-2).
- Saito, A., Hayashi, M., Matsuzawa, T., and Takeshita, H. (2010). Drawing Behaviour in Chimpanzees Compared with Human Development. *霊長類研究 Supplement* 26, 170-170. doi: <https://doi.org/10.14907/primate.26.0.170.0>.
- Sangkloy, P., Burnell, N., Ham, C., and Hays, J. (2016). The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)* 35, 1-12.
- Sarvadevabhatla, R. K., Dwivedi, I., Biswas, A., Manocha, S., and R., V. B. (2017). SketchParse: Towards Rich Descriptions for Poorly Drawn Sketches using Multi-Task Hierarchical Deep Networks. in *Proceedings of the 25th ACM international conference on Multimedia MM '17*. (New York, NY, USA: Association for Computing Machinery), 10-18. doi: <https://doi.org/10.1145/3123266.3123270>.

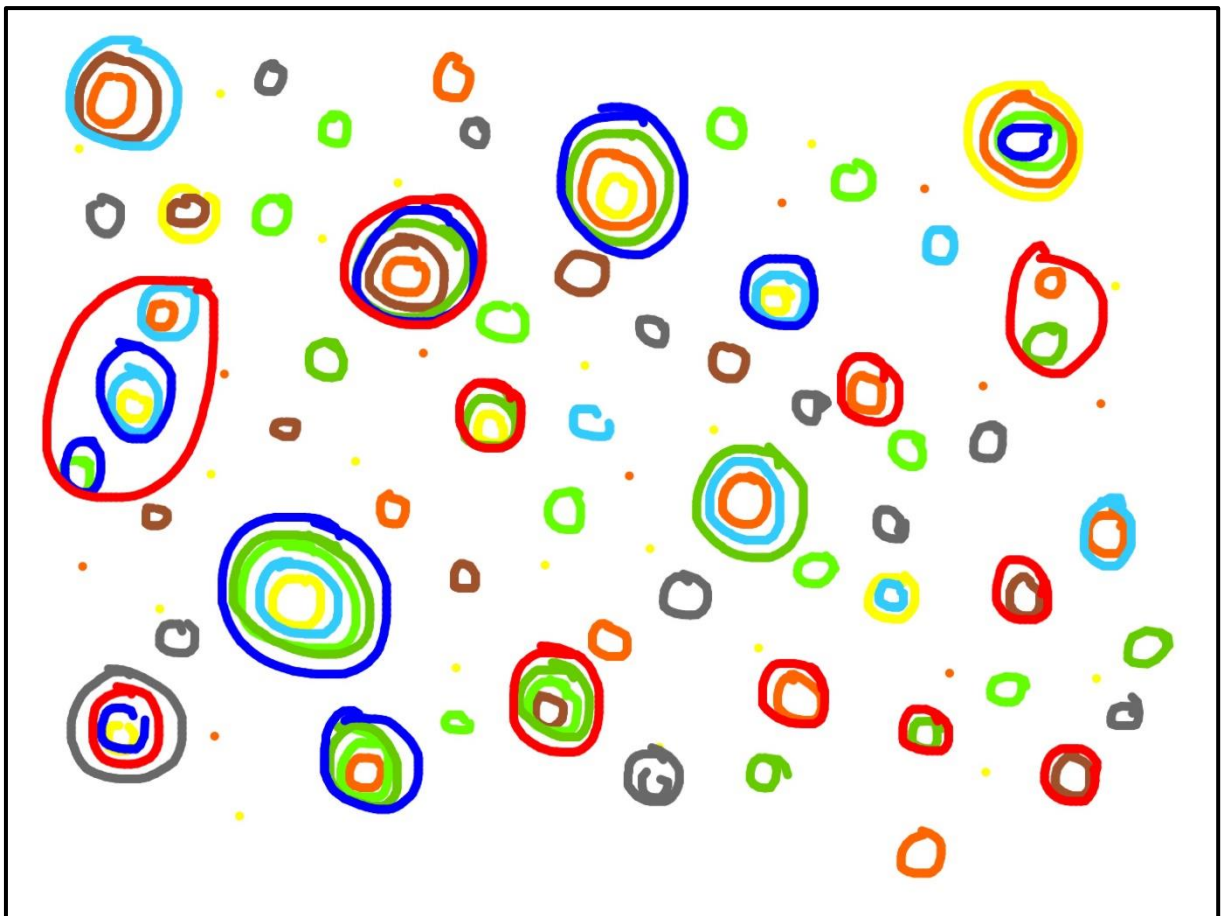
- Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 2673-2681. doi: <https://doi.org/10.1109/78.650093>.
- Simonyan, K., and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*. Available at: <http://arxiv.org/abs/1409.1556> [Accessed November 10, 2021].
- Singer, J., Seeliger, K., and Hebart, M. N. (2020). The representation of object drawings and sketches in deep convolutional neural networks. in Available at: <https://openreview.net/forum?id=wXv6gtWnDO2> [Accessed November 10, 2021].
- Singer, J., Seeliger, K., Kietzmann, T. C., and Hebart, M. N. (2021). From photos to sketches-how humans and deep neural networks process objects across different levels of visual abstraction. doi: <https://doi.org/10.31234/osf.io/xg2uy>.
- Sueur, C., Martinet, L., Beltzung, B., and Pelé, M. (2021). Making drawings speak through mathematical metrics. *arXiv preprint arXiv:2109.02276*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going Deeper With Convolutions. in, 1-9. doi: <http://dx.doi.org/10.1109/CVPR.2015.7298594>.
- Tanaka, M., Tomonaga, M., and Matsuzawa, T. (2003). Finger drawing by infant chimpanzees (*Pan troglodytes*). *Animal Cognition* 6, 245-251. doi: <http://dx.doi.org/10.1007/s10071-003-0198-3>.
- Tharinger, D. J., and Stark, K. D. (1990). A qualitative versus quantitative approach to evaluating the Draw-A-Person and Kinetic Family Drawing: A study of mood- and anxiety-disorder children. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 2, 365. doi: <http://dx.doi.org/10.1037/1040-3590.2.4.365>.
- Theodorus, A., Nauta, M., and Seifert, C. (2020). Evaluating CNN interpretability on sketch classification. in *Twelfth International Conference on Machine Vision (ICMV 2019)* (SPIE), 475-482. doi: <https://doi.org/10.1117/12.2559536>.
- Thomas, G. V., and Silk, A. M. (1990). *An introduction to the psychology of children's drawings*. New York University Press.
- Thomas, G. V., and Tsalimi, A. (1988). Effects of order of drawing head and trunk on their relative sizes in children's human figure drawings. *British Journal of Developmental Psychology* 6, 191-203. doi: <https://doi.org/10.1111/j.2044-835X.1988.tb01093.x>.
- Tirkaz, C., Yanikoglu, B., and Metin Sezgin, T. (2012). Sketched symbol recognition with auto-completion. *Pattern Recognition* 45, 3926-3937. doi: [10.1016/j.patcog.2012.04.026](https://doi.org/10.1016/j.patcog.2012.04.026).
- Tjoa, E., and Guan, C. (2021). A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems* 32, 4793-4813. doi: <https://doi.org/10.1109/TNNLS.2020.3027314>.
- Vaitkevičius, L. (2019). Detection of Scribbles Elements by Image Recognition. Available at: <https://epublications.vu.lt/object/elaba:37989217/> [Accessed November 10, 2021].
- Watanabe, S. (2013). "Animal Aesthetics from the Perspective of Comparative Cognition," in *Emotions of Animals and Humans: Comparative Perspectives The Science of the Mind.*, eds. S. Watanabe and S. Kuczaj (Tokyo: Springer Japan), 129-162. doi: [10.1007/978-4-431-54123-3_7](https://doi.org/10.1007/978-4-431-54123-3_7).
- Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., et al. (2020). Convolutional neural networks for classification of

- Alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis* 63, 101694. doi: <https://doi.org/10.1016/j.media.2020.101694>.
- Wertheimer, M. (1938). "Laws of organization in perceptual forms," in *A source book of Gestalt psychology* (London, England: Kegan Paul, Trench, Trubner & Company), 71-88.
- Widiyanto, S., and Abuhasan, J. W. (2020). Implementation The Convolutional Neural Network Method For Classification The Draw-A-Person Test. in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1-6. doi: [10.1109/ICIC50835.2020.9288651](https://doi.org/10.1109/ICIC50835.2020.9288651).
- Wu, J., Zhang, C., Xue, T., Freeman, B., and Tenenbaum, J. (2016). Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.). Available at: <https://proceedings.neurips.cc/paper/2016/hash/44f683a84163b3523afe57c2e008bc8c-Abstract.html> [Accessed November 10, 2021].
- Wu, T., Sun, W., Li, X., Song, X., and Li, B. (2018a). Towards Interpretable R-CNN by Unfolding Latent Structures. *arXiv:1711.05226 [cs]*. Available at: <http://arxiv.org/abs/1711.05226> [Accessed November 10, 2021].
- Wu, X., Qi, Y., Liu, J., and Yang, J. (2018b). Sketchsegnet: A Rnn Model for Labeling Sketch Strokes. in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1-6. doi: <https://doi.org/10.1109/MLSP.2018.8516988>.
- Xu, D., Read, J. C., Sim, G., and McManus, B. (2009). Experience it, draw it, rate it: capture children's experiences with their drawings. in *Proceedings of the 8th International Conference on Interaction Design and Children*, 266-270. doi: <http://dx.doi.org/10.1145/1551788.1551849>.
- Xu, P., Joshi, C. K., and Bresson, X. (2021). Multigraph transformer for free-hand sketch recognition. *IEEE Transactions on Neural Networks and Learning Systems*.
- Yang, L., Zhuang, J., Fu, H., Wei, X., Zhou, K., and Zheng, Y. (2021). SketchGNN: Semantic Sketch Segmentation with Graph Neural Networks. *ACM Trans. Graph.* 40, 28:1-28:13. doi: <https://doi.org/10.1145/3450284>.
- Yang, L.-C., and Lerch, A. (2020). On the evaluation of generative models in music. *Neural Comput & Applic* 32, 4773-4784. doi: <https://doi.org/10.1007/s00521-018-3849-7>.
- Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding Neural Networks Through Deep Visualization. *arXiv:1506.06579 [cs]*. Available at: <http://arxiv.org/abs/1506.06579> [Accessed November 10, 2021].
- Young-Min, K. (2019). Feature visualization in comic artist classification using deep neural networks. *Journal of Big Data* 6, 56. doi: <https://doi.org/10.1186/s40537-019-0222-3>.
- Yu, Q., Yang, Y., Song, Y.-Z., Xiang, T., and Hospedales, T. (2015). Sketch-a-Net that Beats Humans. doi: <http://dx.doi.org/10.5244/C.29.7>.
- Yue, S. (2017). Imbalanced Malware Images Classification: a CNN based Approach. Available at: <https://arxiv.org/abs/1708.08042v1> [Accessed November 10, 2021].
- Zeiler, M. D., and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. in *Computer Vision - ECCV 2014 Lecture Notes in Computer Science.*, eds. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer International Publishing), 818-833. doi: https://doi.org/10.1007/978-3-319-10590-1_53.

- Zhang, H., Liu, S., Zhang, C., Ren, W., Wang, R., and Cao, X. (2016). SketchNet: Sketch Classification With Web Images. in, 1105-1113. doi: <http://dx.doi.org/10.1109/CVPR.2016.125>.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning Deep Features for Discriminative Localization. in, 2921-2929. doi: <http://dx.doi.org/10.1109/CVPR.2016.319>.
- Zhu, Y., Min, M. R., Kadav, A., and Graf, H. P. (2020). S3VAE: Self-Supervised Sequential VAE for Representation Disentanglement and Data Generation. in, 6538-6547. doi: <http://dx.doi.org/10.1109/CVPR42600.2020.00657>.
- Zou, C., Yu, Q., Du, R., Mo, H., Song, Y.-Z., Xiang, T., et al. (2018). Sketchyscene: Richly-annotated scene sketches. in *Proceedings of the european conference on computer vision (ECCV)*, 421-436.

Chapitre 2

Une première approche mathématique



I. Article 2 : La complexité temporelle du dessin

To draw or not to draw: understanding the temporal organization of drawing behaviour using fractal analyses

Publié dans *Fractals*

Synthèse de l'Article 2

Bien que le dessin soit généralement vu comme un produit, ce dernier est le résultat d'un processus temporel. Afin d'analyser le dessin en tant que processus, il est possible de considérer les séquences temporelles de l'activité du dessin. En d'autres termes des séries binaires prenant la valeur 1 si l'individu dessine à l'instant t et 0 si non. Dans cet article, de telles séries sont étudiées à travers les fractales temporelles, qui permettent de calculer le coefficient de Hurst, un indice permettant d'estimer la complexité temporelle d'une série. Calculé sur les séries temporelles de dessins d'enfants et d'adultes, cet indice a permis de mettre en lumière un comportement plus stéréotypé chez les jeunes enfants que chez les adultes.

To draw or not to draw: understanding the temporal organization of drawing behaviour using fractal analyses

Benjamin Beltzung^a, *Lison Martinet^a, Andrew J. J. MacIntosh^b, Xavier Meyer^{a,b},
Jérôme Hosselet^a, Marie Pelé^c, Cédric Sueur^{a,d}

^a. Université de Strasbourg, CNRS, IPHC UMR 7178, 67000 Strasbourg, France

^b. Primate Research Institute, Kyoto University, Inuyama, Japan

^c. ETHICS EA 7446, Lille Catholic University, Hauts-de-France, France

^d. Institut Universitaire de France, Paris, France

Corresponding author: Lison Martinet

Email : lison.martinet@gmail.com

Address: Institut Pluridisciplinaire Hubert Curien
Département Écologie, Physiologie et Éthologie
23, rue Becquerel
67087 Strasbourg
FRANCE

Phone number: +33 (0)3 88 10 69 30

* Authors equally contributed to this work as first authors.

ABSTRACT

Studies on drawing often focused on spatial aspects of the finished products. Here, the drawing behaviour was studied by analysing its intermittent process, between *drawing* (i.e. marking a surface) and *interruption* (i.e. a pause in the marking gesture). To assess how this intermittence develops with age, we collected finger-drawings on a touchscreen by 185 individuals (children and adults). We measured the temporal structure of each drawing sequence to determine its complexity. To do this, we applied temporal fractal estimators to each drawing time series before combining them in a Principal Component Analysis procedure. The youngest children (3 years-old) drew in a more stereotypical way with long-range dependence detected in their alternations between states. Among older children and adults, the complexity of drawing sequences increased showing a less predictable behaviour as their drawings become more detailed and figurative. This study improves our understanding of the temporal aspects of drawing behaviour, and contributes to an objective understanding of its ontogeny.

Keywords: temporal complexity, drawing intermittence, marking gesture, anthropology, evolution, mathematics, *Homo sapiens*

1. INTRODUCTION

The expression of artistic behaviour is predominant in *Homo sapiens*, even if the time allocated to such activities tends to decrease with age. A good example of such declining interest across the lifetime can be seen in drawing. Prevalent among children, drawing persists as a recreational activity only among a minority of adults, and as a professional activity among a select few. Drawing develops from internal representativeness (i.e. a drawing representative only from the perspective of the child who produces the drawing) to external representativeness (i.e. a drawing interpretable from the perspective of an independent observer) around the age of 3-4 years (Freeman, 1993; Martinet et al., 2021). Drawing behaviour has been studied in different research fields, especially in psychology, e.g. to characterise its ontogeny (Luquet, 1927; Willats, 2005), its representational development (Adi-Japha et al., 1998; Cherney et al., 2006), the use of colour (Wright & Black, 2013), the comprehension of realism (Jolley et al., 2000) or the influence and role of gender (Cherney et al., 2006; Turgeon, 2008). Such studies have provided critical information about drawing behaviour, despite the fact that the analyses used are often qualitative and subject to biases.

Adi-Japha et al (1998) showed that young children, even if not capable of attributing meanings to the entirety of their drawing, can make sense of broken lines which seem to be more descriptive than curved ones. Although relevant, this method remains subjective because researchers directly question children about what their drawings represent. Moreover, very young children (toddlers) are unable to express themselves with respect to their drawings. Furthermore, the majority of drawing studies focus on spatial measures, whereas the temporal aspects that might also be important in revealing cognitive and decision-making processes remain

underexplored. Recent developments in mathematical analyses in animal behaviour might bring new perspectives to the study of such processes.

Indeed, new analytical approaches from statistical physics are now being developed to study individual behaviour, such as fractal analyses. An object is considered fractal when any of its parts examined separately resemble its overall structure at any magnification scale, either in the spatial or temporal domain (Mandelbrot, 1977). A good example of a natural fractal object is Romanesco broccoli, but many other natural systems have fractal structure in the spatial domain, such as river networks (Rinaldo et al., 1993) and human lung architecture (Nelson et al., 1990). Fractal structure is also found in the temporal domain, in processes such as human breathing cycle dynamics (Peng et al., 2002) or human sleep EEG (Weiss et al., 2009). In the behavioural sciences, recent studies have established that fractal analyses of spatial and temporal patterns can lead to a better understanding of animal and human behaviour (MacIntosh, 2014; Rutherford et al., 2003).

On one hand, the study of movement behaviour through spatial fractal analyses have allowed comparisons between observed movement patterns and theoretically optimal foraging patterns (Sims et al., 2008; Viswanathan et al., 1999). We previously applied this approach in examining the efficiency of the drawing trajectory, defined as the correct reading of the drawing with minimal detail in chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*). Analogous to the trajectory of an animal in its environment, we wanted to know if the efficiency of drawings made on a screen differed between humans and chimpanzees, and with age in humans. Results show that the drawing spatial index was lowest in chimpanzees, increased and reached its maximum between 5-year-old and 10-year-old children, and decreased in adults,

whose drawing efficiency was reduced by the addition of details (Martinet et al., 2021).

On the other hand, temporal fractal analyses have shown that an animal's physiological condition (e.g. health status or degree of physiological stress) affects its sequences of behaviour (Alados, 1996). The concept of fractal time reflects the degree to which current behavioural states depend not only on states immediately preceding them in the sequence but also on those that occur much earlier in the sequence (so-called long-range correlation or long-memory processes) (Delignières et al., 2005). Like spatial fractal analyses, temporal fractal analyses also suggest an optimal structure in behaviour time series, within which a normally functioning individual should fall (MacIntosh, 2014). Some environmental pressures or physiological impairments can lead to a loss of complexity in the behavioural sequences of an individual, associated with increases in periodicity or long-range dependence (i.e. greater stereotypy) (MacIntosh et al., 2011; Maria et al., 2004; Meyer et al., 2020). Other factors may in contrast lead to increased complexity, i.e. reduced long-range dependence and increased behavioural stochasticity (Rutherford et al., 2003). To summarise or simplify the concept, certain complexity signatures in behaviour sequences equate to increased stability and/or adaptability, which is the case for some physiological processes as well as sequences of animal behaviour, where it can be an indicator of well-being (Alados, 1996; Maria et al., 2004). By allowing more detailed insight into behavioural sequences, temporal fractal analyses open up new research perspectives for diverse fields of study.

In humans, behavioural sequences are complex and structured in time by multiple factors (e.g., environmental, psychological, physiological), which may limit randomness in the occurrence of behaviour. Several behaviours have been described

as fractal, such as patterns of physical activity (Paraschiv-Ionescu et al., 2008) or short-message communication in online communities (Rybski et al., 2009). Considering drawing and handwriting, both pacing and grip force showed a fractal dimension (Fernandes & Chau, 2008). However, the latter study was restricted to an analysis of young adults drawing circles in synchrony with a metronome, so these results cannot be generalised.

In the present study, we sought to know whether temporal fractal analyses could be applied to drawing data from humans of all ages to learn more about its expression and its ontogenetic processes. Drawing behaviour is an intermittent process, composed of two main states, including *drawing* (i.e. marking a surface) and *interruption* (i.e. a break in the marking gesture). The temporal structure of each drawing sequence (i.e. sequence memory) can be measured to determine its degree of complexity (MacIntosh et al., 2011). To assess how drawing intermittence develops with age, we collected drawings from 185 individuals of different ages, including children as well as both naive and expert adults. We hypothesized that drawing sequences would be simpler in young children (aged 3 to 5 years), with greater periodicity and long-range dependence in their alternations between states. Indeed, young children who have not yet entered the representative phase of drawing may be more interested in gestures and engaged more in play than in drawing. On the contrary, complexity in drawing sequences may increase among older children and adults, for whom drawings become more and more detailed and representative.

To accomplish our aims, we applied multiple temporal fractal estimators to study long-memory processes in drawing time series. A first challenge was to link these few estimators to drawing behaviour, as we are not aware of any previous studies on this topic. A second challenge was to assess the effectiveness of these fractal

estimators, as there is little consensus as to which are most appropriate for a given set of data (Stadnitski, 2012; Stroe-Kunold et al., 2009). Therefore, we first present a robust analytical approach to performing these fractal analyses using our drawing dataset to decrease errors and biases. We then introduce the analysis of complementary indices such as the duration of drawing behaviour in relation to the duration of the drawing session, and the number of swings between drawing and non-drawing states during the drawing session. We discuss our results in the light of biological processes.

2. MATERIAL AND METHODS

2.1 Ethics

We followed the ethical guidelines of our research institutions and obtained ethical approval from the University of Strasbourg Research Ethics Committee (Unistra/CER/2019-11). Drawings were anonymously collected. The contribution of all participants was voluntary and subject to parental consent for children.

2.2 Participants

One hundred and forty-four children (67 girls and 77 boys) and forty-one adults (21 women and 20 men) participated in this study (Table 1). We worked with children from kindergarten and primary school in Strasbourg, France. We collected their drawings over two consecutive years (in 2018 for kindergarten children and in 2019 for primary school children). For this reason, 6-year-old children could not be tested in 2019 since they had already participated in the study the year before. Adult participants were between 21 and 60 years old. In addition to a general age effect, we also tested the effect of experience in our adult participants. Among researchers and students at the authors' research institute, twenty were considered naive (naive adults: 30.8 ± 10.54 years old) as they had never taken drawing lessons and did not

draw as a hobby. The other twenty adults were classified as experts and included art school students and professional illustrators (expert adults: 30.4 ± 11.12 years old).

Table 1. Groups and gender of human participants.

Participants groups		Gender	Number of participants
Children	3-year-olds	girls	n=5
		boys	n=15
	4-year-olds	girls	n=10
		boys	n=10
	5-year-olds	girls	n=10
		boys	n=10
	7-year-olds	girls	n=12
		boys	n=11
	8-year-olds	girls	n=9
		boys	n=9

2.3 Experimental design

2.3.1. Habituation phase

Children and adults were invited to try the device, a touchscreen tablet (iPad Pro, 13-Inch, version 11.2.2) by drawing on it with their fingers. Participants could use 10 different colours displayed on the bottom of the screen and selected one of them by clicking on it. Adults were tested immediately after habituation but, to avoid overstimulation, children were habituated the day before their respective tests.

2.3.2. Testing phase

Children were tested individually at school, either in their own classrooms (for 3-year-olds) or in the staffroom (for older children). During the test, the experimenter (LM or MP) stayed with the child but kept their distance so as not to influence their behaviour. Adults participated individually, either alone in a room at the research institute (for naive subjects) or at the art school (for experts). We used a video camera to record the hand movements of each participant while drawing, in case we needed to control for any contextual issues arising during the session (e.g. disruption of the drawing session, unintentional points or lines made by the experimenter at the end of the session, *etc.*). No time limits were imposed during the study.

Participants were tested under two conditions to assess potential differences between a non-specific task (*free drawing*) and a specific task (drawing a *self-portrait*). Under the *free condition*, subjects were asked to draw whatever they wanted, with no further instructions. Under the *self-portrait condition*, subjects were asked to draw themselves. In each participant category, we randomly assigned subjects such that half began the test with the *free condition* instructions while the other half began with the *self-portrait condition*. To avoid overstimulation and a lack of concentration, none of the children participated in both conditions on the same day.

2.3.3. Data collection

Three hundred and sixty-nine drawings were collected (one drawing by a naive adult was deleted by mistake). Since there were no imposed time limits, so as to not constrain the creativity of each participant, the drawing duration was different for each person and ranged from 17 seconds to 923 seconds (mean \pm s.d. = 250.5 \pm 189.2). When the individual was drawing, a triplet (x,y,t) was recorded every 17 milliseconds

on average (resolution of point recording), where (x, y) is the position of the finger on the touchscreen and t the time. For two consecutive points (x_i, y_i, t_i) and $(x_{i+1}, y_{i+1}, t_{i+1})$, the time interval in milliseconds is $t_{i+1} - t_i$. If $t_{i+1} - t_i < 100$, we considered it drawing time. If $t_{i+1} - t_i > 100$, we considered it non-drawing time; i.e., the individual must raise their finger long enough for this stop to be voluntary (Tanaka et al., 2003). This succession of time intervals was then transformed into a binary series of -1 (not drawing) and 1 (drawing), where the duration between two consecutive points was set at 1 millisecond. Note that analyses with other resolutions (10 milliseconds and 17 milliseconds) were performed for comparison and gave similar results (see Appendix A. 1). The resultant time series can be represented as a barcode as shown in Fig. 1. The width of each black band corresponds to a drawing duration, while a non-drawing duration is equivalent to the width of a white band. Each video was watched several times to remove moments during which participants were distracted by something else (Longstaff & Heath, 1999) (simulations showed that this did not mathematically impact the results, see Appendix Box1) . This happened a few times at school during drawing sessions with children. This concerns only 22 drawings in our final analyses (i.e. 6.4% of the dataset) where 1 to maximum 2 durations were removed, representing $12.2 \pm 7.1\%$ of the final sequence (these were generally times exceeding 15 seconds on average and caused by a child entering the room, the bell for recess ringing, etc).

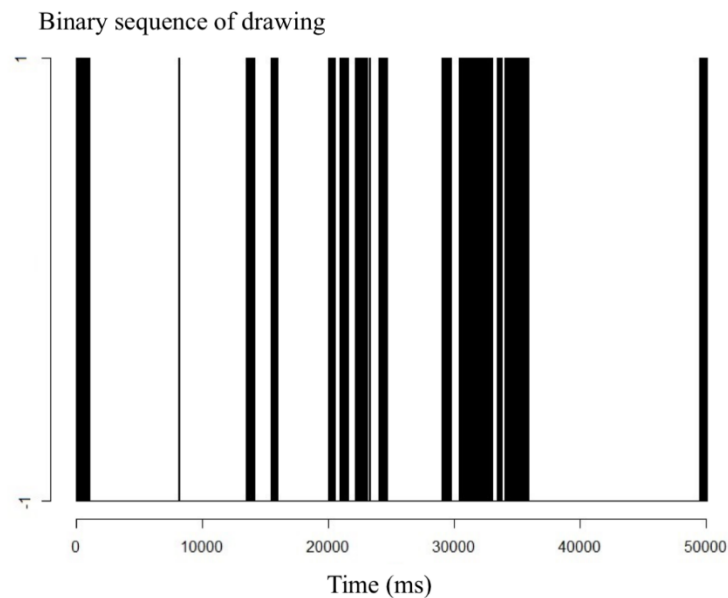


Figure 1. Example of a binary sequence of intermittent drawing, denoted 1 for drawing and -1 for non-drawing behaviour. Black bars reflect the durations of the drawing state while white bars reflect the durations of the non-drawing state.

2.4 Analyses

2.4.1. Type of signal

Fractal characteristics or long memory processes can be measured via different algorithms, each having its own statistical parameter. Here, the difficulty arises from the fact that, for each parameter, numerous estimators have been defined, but the effectiveness of each is still debated in the literature (Stadnitski, 2012; Stadnytska et al., 2010). Studies often focus on one or few estimators without a rigorous reason for comparing them. As a consequence, there is no simple and systematic way to estimate the fractal process, which often results in errors or misleading conclusions (Karagiannis et al., 2006).

The most widespread way to assess and quantify long memory processes in temporal sequences is estimation of the Hurst exponent (H). This exponent is a measure of the correlation among signal components in a time series (Cannon et al., 1997; Stroekunold et al., 2009). Within this framework, behavioural sequences can reflect three

mutually-exclusive scenarios: (1) *persistence* ($H > 0.5$) occurs when positive long-range autocorrelation exists, such that blocks of certain behaviours (e.g. drawing) are likely to be followed by blocks of similar duration in succession; (2) *anti-persistence* ($H < 0.5$) occurs when negative long-range autocorrelation exists, such that blocks of behaviour are likely to be followed by blocks of divergent duration; (3) *white noise* ($H = 0.5$) occurs when no sequence memory exists, i.e., the sequence is random or contains only short-range autocorrelation (Delignières et al., 2005).

Methods for H estimation differ depending on the signal class of the original sequence, which can be either *Fractional Gaussian noise* (fGn) or *fractional Brownian motion* (fBm) (Mandelbrot & Van Ness, 1968). fGn is stationary with constant variance and mean whereas fBm is nonstationary, even if both signals are theoretically linked: differencing fBm produces fGn and integrating fGn creates fBm (Stadnytska et al., 2010). The same original sequence expressed as one or the other signal class will be characterized by the same Hurst exponent (Fig. 2). However, before estimating a behaviour sequence's H exponent, it is essential to first define its signal class, which can be done through different methods (Cannon et al., 1997). The so-called 'temporal' methods are those which do not require prior transformation of the data and identify statistical dependence in elements of the time series (Eke et al., 2002). Frequency-based methods, on the other hand, are based on transformation of the time series, for example by considering the periodogram (i.e. the spectral density of a signal). In this article both types of methods are used with the aim of comparison and to combine the different approaches for a more robust investigation. The four methods we use do not allow to calculate H directly but indirectly through parameters that are linked to it.

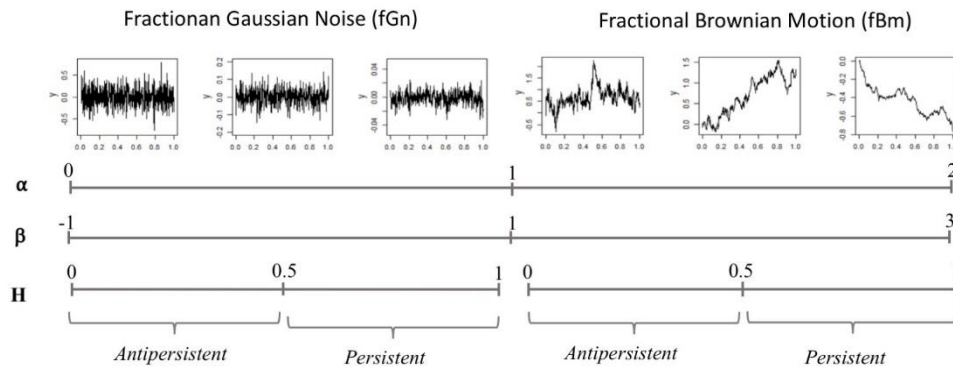


Figure 2. Illustration of the fGn/fBm continuum. Values of the scaling exponent α (from detrended fluctuation analysis: DFA) and the index β (from power spectral analysis) are depicted in relation to the Hurst exponent, H . This representation is largely inspired by (Marmelat et al., 2012).

2.4.2. Temporal methods

Detrended Fluctuation Analysis (DFA)

To investigate long-memory processes in the sequential distribution of drawing and non-drawing durations, we employed Detrended Fluctuation Analysis (DFA) (Peng et al., 1995) which is among the most used to study binary sequences of animal behaviour (MacIntosh et al., 2013; Meyer et al., 2020; Rutherford et al., 2003). It is also a robust estimator of the Hurst exponent (Cannon et al., 1997; Eke et al., 2002). DFA calculates a scaling exponent (α_{DFA}) corresponding to the slope of the line on a log-log plot of the average fluctuation at each box size (given by the three steps, from equation 1 to equation 3 below). A lower α_{DFA} reflects greater stochasticity and reduced long range memory (Fig. 2) (Meyer et al., 2017). If a linear relation exists, it indicates the presence of scale invariance.

To summarize the application of DFA to our data, a binary sequence of a drawing (Fig. 3a) is cumulatively summed such that

$$y(t) = \sum_{i=1}^t x(i) \quad (1)$$

Where $y(t)$ is the cumulative sum and $x(i)$ is the binary sequence at each time step (1ms intervals).

The cumulative sum is then divided into non-overlapping boxes of length n (Fig. 3a) and a least-squares regression line is fit to each box $y_n(t)$ to remove local trend and it is repeated over all box sizes, then the fluctuation is calculated as follows (Fig. 3b)

$$F(n) = \sqrt{1/N \sum_{i=1}^N (y_n(t) - \hat{y}_n(t))^2} \quad (2)$$

Where N is the length of the series, $\hat{y}_n(t)$ is the regression estimate for $y_n(t)$ at each box size n and $F(n)$ is the fluctuation of the modified root-mean-square equation across all scales available ($2^2, 2^3, 2^4 \dots 2^m$) where m is the largest scale examined, such as $2^m < N/2$ (rounded down to the nearest whole number) (Constantine & Percival, 2007; Stadnytska et al., 2010; Stroe-Kunold et al., 2009). We then obtained the following relationship

$$F(n) \sim n^\alpha \quad (3)$$

This method makes it possible to know whether a series is a fBm or a fGn, since it can be used on both (Seuront, 2009; Stadnytska et al., 2010; Stroe-Kunold et al., 2009). The α exponent can be interpreted as follows:

If $\alpha_{DFA} > 1$, the original signal is of the signal class fBm and $H = \alpha_{DFA} - 1$

If $\alpha_{DFA} < 1$, the original signal is of the signal class fGn and $H = \alpha_{DFA}$

α_{DFA} in the range of 0.5-1 and 1.5-2 reflects persistence while α_{DFA} in the range of 0-0.5 and 1-1.5 reflects anti-persistence for fGn and fBm, respectively (Fig. 2). Values of 0.5 and 1.5 reflect Gaussian (white) noise and Brownian motion, respectively.

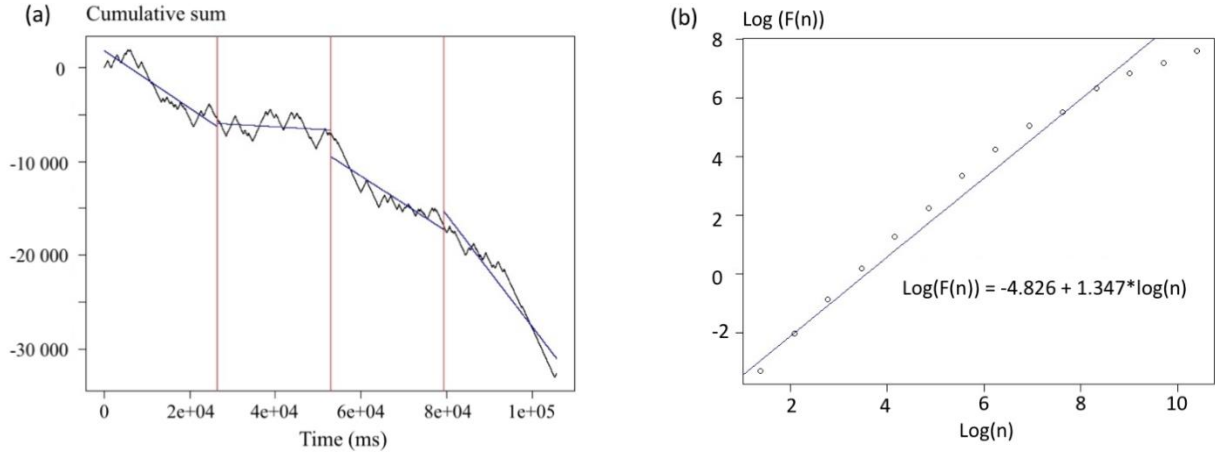


Figure 3. (a) Integration of a drawing sequence. Vertical lines give an example of dividing the sequence into $N/n = 4$ boxes. Lines in each box correspond to the polynomial regression. (b) Bi-logarithmic plot of the statistic $F(n)$ against the length of the time intervals n . The regression line allows calculation of the scaling exponent α , which is its slope (1.347 in this case).

Hurst Absolute Value (HAV)

The HAV method is similar to DFA but does not integrate the time series before analyses so that H is calculated from the original binary sequences of drawing and non-drawing durations. We included the HAV method because it is able to capture the self-similarity parameter in time series data where DFA fails to do so (Mercik et al., 2003). This method only works on fGn, so any application of HAV on fBm signals must use their increments (i.e. the corresponding fGn). A time series of class fGn of length N is divided into smaller boxes of length n , denoted as $x^{(n)}$, and the first absolute moment is obtained as follows:

$$\delta^{(n)} = 1/(N/n) \sum_{k=1}^{N/n} |x^{(n)}(k) - \bar{x}| \quad (4)$$

This is reiterated for the different window size n with the variance δ varying as follows:

$$\delta^{(n)} = n^{H_{AV}-1} \quad (5)$$

Where H_{AV} is the scaling exponent.

Scaled Windowed Variance (SWV)

Since we only had fBm series, the SWV method was selected as it a good estimator for this type of signal (Delignieres et al., 2006). The fBm series $x(t)$ is divided into non-overlapping boxes of length n . In each box, a bridge detrending method is used to remove the trend, this is what is recommended for series of more than 2^{12} points (Cannon et al., 1997), which is our case here. Then the standard deviation (SD) is calculated in each box such that

$$SD(i) = \sqrt{\sum_{t=1}^n [x(t) - \bar{x}_i]^2 / n - 1} \quad (6)$$

With \bar{x}_i being the average in the box i . Then, the average standard deviation (\overline{SD}) of all boxes of length n is calculated such that

$$\overline{SD(n)} = 1/m \sum_{i=1}^m SD(i) \quad (7)$$

Where m is the number of boxes of length n .

This step is then repeated over all possible box lengths and $\overline{SD(n)}$ is related to n by a power law

$$\overline{SD(n)} \propto n^{H_{swv}} \quad (8)$$

The $\log(\overline{SD(n)})$ is plotted against $\log(n)$ and the slope of the regression is the estimated Hurst coefficient H_{swv} (Cannon et al., 1997; Delignières et al., 2005).

2.4.3. Frequency-based method

Power Spectral Density analyses (PSD)

Like DFA, this method is widely used to determine the signal class of time series data. It is included here for comparison, because of the widespread use of PSD in standard time series analysis, and the added diversity it affords us as an index based

on frequency. In the frequency domain, the fractal character is expressed through the following power law:

$$E(f) \propto 1^{-\beta}/f \quad (9)$$

where $E(f)$ is the squared amplitude for the corresponding f frequency. The β exponent is obtained by calculating the negative slope ($-\beta$) of the regression between $\log(E(f))$ and $\log(f)$ (Fig. 4). β values range between -1 and 3, with β between -1 and 1 reflecting fGn and β between 1 and 3 reflecting fBm (Delignières et al., 2005; Eke et al., 2000). β is linked to H such that:

$$H = (\beta + 1)/2 \text{ for fGn}$$

$$H = \beta - 1/2 \text{ for fBm}$$

In the present study, the high frequencies ($1/8 < f < 1/2$) are excluded. This method provides more reliable estimates of β and is known as $lowPSD$ (Stadnitski, 2012).

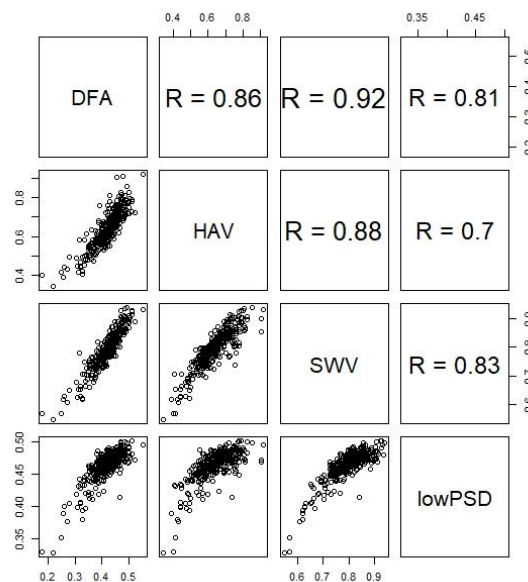


Figure 4. Bi-logarithmic power spectrum of a time series with regression slope fitted to obtain the β index.

2.4.4 Methods implementation

We performed DFA using the package *fractal* (Constantine & Percival, 2007) in R. For the Power Spectral Density, we based our analyses on the script found in Stadnitski et al. (2012). As the Scaled Windowed Variance and Hurst Absolute Value methods are not currently implemented in R, we coded and tested our own algorithms through Monte-Carlo simulation. We simulated, using the *somebm* package (Huang, 2013), fBm of different lengths and known H . In these simulations, the lengths of the series were defined as powers of 2, from 2^{11} (= 2,048 points) to 2^{15} (= 32,768 points). For each possible length, 100 fBm were generated with $H = 0.2$, $H = 0.4$, $H = 0.6$ and $H = 0.8$. More than 2000 series were generated in total and the errors, $H - \hat{H}$, are shown in the Appendix (A. 2 and A. 3). The drawing sequences obtained in this study were much longer (mean \pm s.d. = 240,832ms \pm 185,561ms) than these simulated series and therefore will not be a limiting factor in the estimates. Since the HAV method only works on fGn, the simulated series used are those of the respective fBm increments. These simulations have shown that the larger the size of the series, the closer the errors are to 0, proving the efficiency of these analyses.

2.4.5 H estimation

The Hurst coefficient H was estimated using the four methods described above. The binary time series studied all being fBm, it was necessary to consider series consisting of the increments of these fBm (i.e. the corresponding fGn) to be able to apply certain methods such as HAV. As mentioned earlier, each of these methods does not always directly estimate the Hurst coefficient, H , but the latter's relationship with the calculated exponent is relatively simple (Fig. 2). Once the four methods were applied to each series, a Pearson correlation matrix was calculated.

2.4.6 Additional temporal indices

Both of the indices described below were extracted from the complete binary sequences.

Proportion of drawing in the sequence

We first calculated the number of 1's in the sequence. Since each binary sequence had a different length, we divided the sum of 1 by the total sequence length to come up with a drawing proportion for comparison across sequences.

Rate of state changes

The number of state changes corresponds to the number of times a participant changed their behaviour between drawing and interruption during the drawing session. For each sequence, the number of state changes was divided by the total sequence length and multiplied by 1000 to obtain a rate of change per second.

2.4.7 Statistical analysis

Statistical analyses were conducted in R (version 3.6.2). Since we used four methods to estimate H , we obtained 4 estimated values of this index for each time series to use in our statistical analyses. Some previous studies have advised averaging the different estimates, but the choice of which to use and the overall number of estimators used has varied (Eke et al., 2000; Seuront, 2009). The four estimates of H were found to be highly correlated with each other. Considering that averaging variables whose variance may be different can remove information, we chose to perform a PCA. In our case, the method was used to reduce the number of estimators used into one that retained as much of the information as possible. The new resultant variable was thus constructed as a linear combination of the original variables, allowing for a synthesis of our variables into a single index (Berni et al., 2011). After

using the Kaiser-Guttman criterion to select the number of axes, only values of the first principal component from the PCA - a proxy for the Hurst exponent that we term the '*Hurst axis*' - were used and set as a response variable in our statistical models. In our case, results obtained following this methodology with those resulting from averaging the estimates were equivalent (see Results). We decided to retain the PCA, as it makes it possible to create a new variable (i.e. an index) which is an optimally weighted combination of correlated estimators of *H*.

A correlation of -0.43 was observed between the sequence length and the Hurst axis. Indeed, there were considerable differences in the lengths of our drawing sequences, which averaged approximately 4 minutes and ranged between approximately 17 seconds and 15 minutes (mean \pm s.d. =240832 \pm 185561ms; range=16848-908250ms). We were not interested in this correlation but needed to account for it when testing relationships between *H* and our other variables of interest, such as the *gender*, the *group* (age) or the *condition* under which the drawing was made. For this reason, the four estimated coefficients of *H* were recalculated on the first 50,000 points (i.e. the first 50 seconds of the drawing), 100,000 points, 150,000 points and 200,000 points. DFA coefficients based on the first 50,000 points in the sequence were correlated with those based on these other three lengths (Appendix, A. 4), as well as with those from the whole sequence at more than 54%, meaning that the information contained in the first 50,000 points is a good threshold compared to that contained in the entire sequence. Therefore, we analysed only the first 50,000 points of each sequence and excluded all sequences less than 50,000 points in length (6.5% of the drawings collected).

Given the above criteria, we analysed 346 drawings out of the initial 369 we obtained. After recalculating the indices on the first 50,000 points for each

sequence, the PCA procedure was redone. We then determined whether *group*, *gender* or *condition* were associated with variation of the “Hurst axis” by constructing a Generalized Linear Mixed Model with Gaussian error structure (*nlme* package (Pinheiro et al., 2006)). Since each participant produced two drawings, *individual identity* was added as a random factor. Residual normality was graphically verified. The normality of the residuals of the random effect was graphically validated for each group. Since heteroscedasticity was detected across groups in the original model, we added a covariance structure (*VarIdent*, adapted to the categorical variables) to allow the variance of the residuals to change according to group. The full model included all possible variables and first-order interactions between *group*, *gender*, and *condition*. We proceeded to model selection using a dredge function based on the lowest Aikake’s Information Criterion (package *MuMIn*; $\Delta AIC > 2$ (Barton, 2009)). Paired comparisons were made using the *lsmeans* package to compare different age groups in pairs.

Concerning the two additional metrics, the proportion of drawing in the sequence and the rate of state changes, measures have been done on the binary sequences. For the number of states changes, we normalized the data using a Box-Cox transformation and ran a Linear Mixed Model (package *nlme*) containing the variables *group*, *gender* and *condition* with *individual identity* added as a random factor. Again, model selection was carried out using the dredge function (package *MuMIn*) and we chose the model with the lowest AIC. Residual normality was graphically verified, but since heteroscedasticity was detected, we added a covariance structure to allow for a difference in the residuals variance between groups (*VarIdent*). For the proportion of drawing, we used a Linear Mixed Model (package *nlme*) too. The

conditions of application (residual normality and homogeneity of variance) were graphically verified. The alpha level for all statistical analyses was set at 0.05.

3. RESULTS

3.1 Type of signal

Examination of α through DFA (mean \pm sd = 1.419 ± 0.0490) shows that the original binary sequences were characteristic of fBm. This result was confirmed by the ^{low}PSD method which estimates of β were greater than 1 (mean \pm sd = 1.927 ± 0.0472).

3.2 Estimates of H

The Hurst exponent was estimated with the four methods (Appendix A. 5). The binary time series studied being all fBm, we considered series resulting from the increments of these fBm. Each of the methods is positively correlated to the others, an expected result since they estimate the same coefficient (Fig. 5).

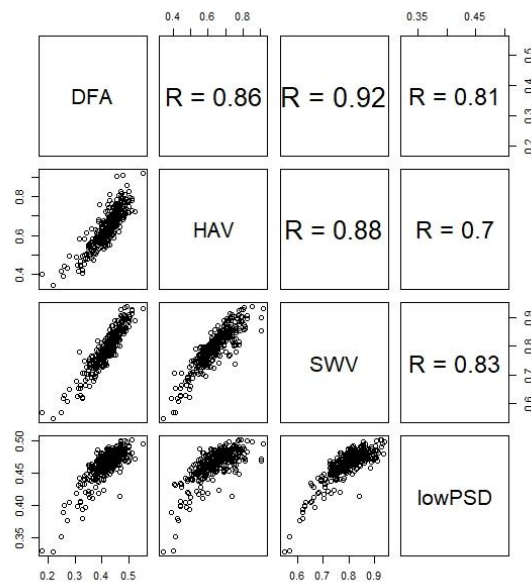


Figure 5. Pearson correlation matrix indicated the correlation values between the four methods used to estimate the Hurst exponent H .

3.3 Combination of the estimators through a Principal Component Analysis (PCA)

Since the PCA was standardized, only the axes whose inertia was strictly greater than 1 were kept, which was equivalent to keeping only axis 1 in our data set, explaining 87.68% of the variance (Correlation circle available in Appendix A. 6). All H estimators loaded positively into the first principal component, i.e. the Hurst axis (the loadings for DFA, HAV, SWV, and ^{low}PSD were 0.513, 0.491, 0.518, and 0.475, respectively).

3.4 Variation in the Hurst axis according to the group

The selected model was the one that contained only the variable *group* ($df = 8$, $\chi^2 = 21.434$, $p = 0.006$; Fig. 6). Paired comparisons indicated that two significant differences emerged: 3-year-olds had a higher value along the Hurst axis than novice ($p = 0.0085$, $t = 3.706$) and expert ($p = 0.0148$, $t = 3.540$) adults. Neither gender nor condition was associated with variation in the Hurst axis.

Furthermore, values of the Hurst axis were highly correlated with the average value across our 4 estimators ($r = 0.993$), and the same statistical model, containing only group as an independent variable, best explained variation in this averaged metric (Appendix Table 1).

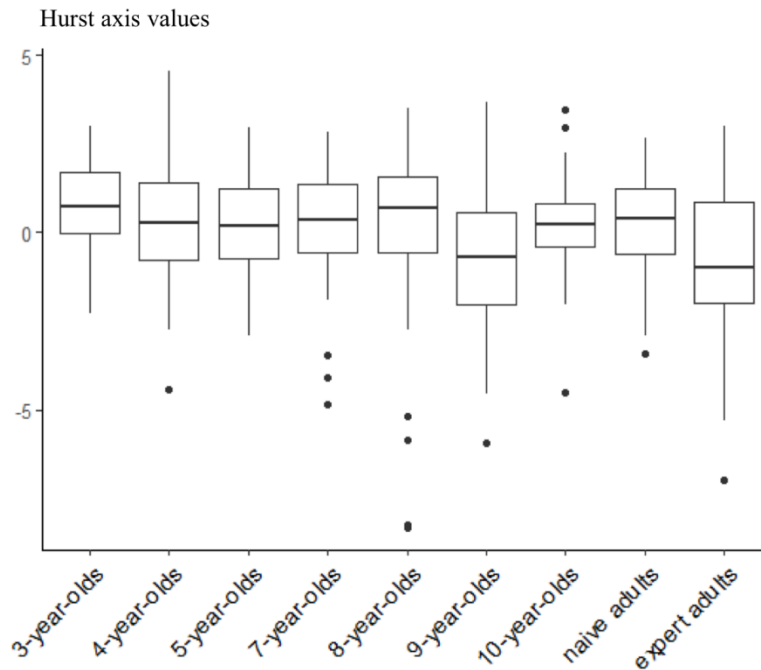


Figure 6. Boxplots of the Hurst axis values for each group. Each boxplot depicts the median (bold bar), 25-75% quartiles (box) and outliers (points).

3.5 Additional temporal indices

3.5.1 Proportion of drawing in the sequence

The selection of models led us to choose the model containing the variables *group* ($df = 8$, $\chi^2 = 65.559$, $p < 0.0001$) and *condition* ($df = 1$, $\chi^2 = 13.042$, $p = 0.0003$). Naive adults showed a proportion of drawing significantly lower than that of all other participants ($p < 0.005$, $t < -3.845$; Fig.7). In addition, the proportion of time spent drawing as higher in the *free* condition compared to the *self-portrait* condition ($p = 0.0004$, $t = 3.611$; Fig. 7).

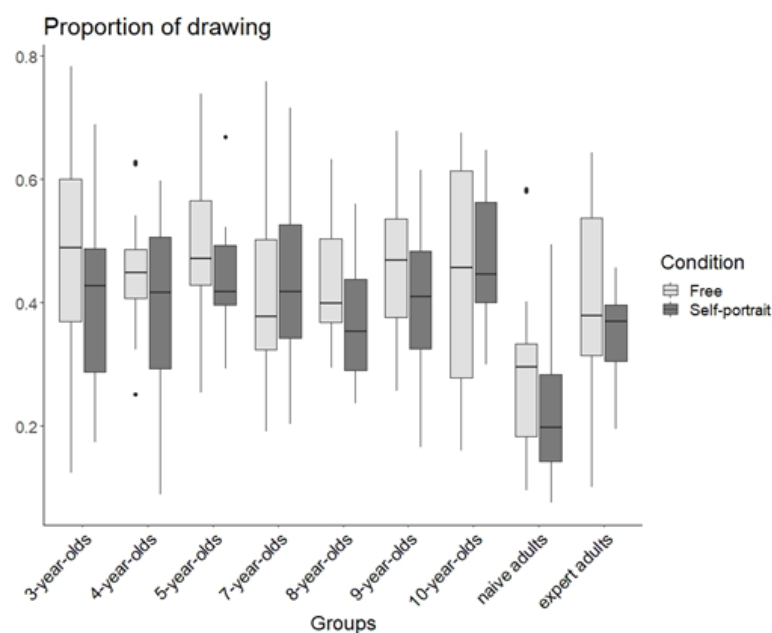


Figure 7. Boxplot of the proportion of drawing for each group and for both conditions. Each boxplot depicts the median (bold bar), 25-75% quartiles (box) and outliers (points).

3.5.2 Rate of state changes

The selected model was the one containing only the variable *group* ($df = 8$, $\chi^2 = 32.607$, $p < 0.0001$). The 3-years-old children alternate significantly less between drawing and interrupting behaviours than 5-year-olds ($p = 0.0310$, $t = -3.303$), 8-year-olds ($p = 0.0262$, $t = -3.359$) and adults, both naive ($p = 0.0011$, $t = -4.261$) and expert ($p = 0.0009$, $t = -4.303$) (Fig. 8). No other significant effects, gender or condition, were found.

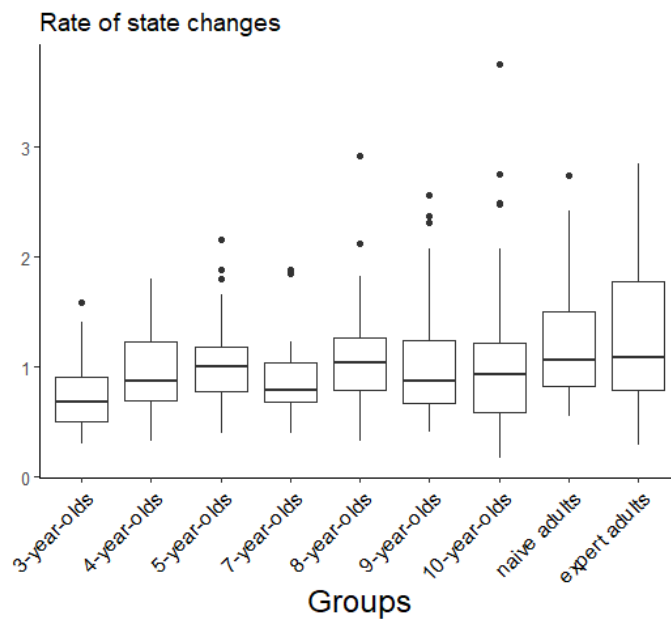


Figure 8. Boxplot of the rate of state changes (number of state changes per second) for each group. Each boxplot depicts the median (bold bar), 25-75% quartiles (box) and outliers (points).

4. DISCUSSION

With this work, we wanted to know whether temporal fractal analysis could be applied to the study of drawing and if it could provide insight about it, as has been done in other fields (MacIntosh, 2014) but not yet - to our knowledge - to understand intermittence in drawing. Specifically, we tested whether the temporal dynamics of drawing varies according to the age, gender or instruction under which the drawing was carried out. In a previous study applying spatial fractal analyses in the field of drawing behaviour, we had access to the efficiency of the drawing trajectory, defined as the correct reading of the drawing with minimal detail. We found that this index increased during childhood and reached its maximum in children between 5 and 10 years of age, before decreasing in adults due to the addition of greater detail (Martinet et al., 2021). With the present study, we show that it is possible to apply temporal fractal analyses to the behavioural sequence of a drawing and we highlight first results.

We found a difference between the youngest (3-year-old children) and oldest (adults) participants, meaning that these two groups do not draw with the same patterns of alternation between drawing and non-drawing states. In other words, the process of drawing shows different degrees of temporal complexity. Young children showed the highest values of the Hurst axis, suggesting that they exhibit more stereotypical and therefore less complex temporal patterns in their drawing behaviour. On the contrary, temporal sequences of drawings were found to be more complex, meaning less predictable, in both naive and expert adults. In relation to this, the rates of state change we observed demonstrate that 3-year-olds performed significantly fewer alternations between drawing and drawing interruptions than all adults, 8-year-olds and 5-year-olds. Said differently, the drawings of the youngest children are composed of fewer strokes per unit of time spent on the activity.

In general, previous work has shown that 3-year-old children draw for shorter periods than older participants, get bored faster and may be more motivated by the simple motor pleasure of using the tablet rather than recognizing it as a real drawing support tool (Martinet et al., 2021). Many young children first tried each available colour, one by one, which might have induced a certain stereotypy in alternations between drawing and interruption, leading to high Hurst axis values. Their drawings were comprised of what could be called scribbles, as they were not figurative, not representative, at least to the eye of an external observer (Martinet et al., 2021).

However, 3-year-old children did not stand out in terms of the proportion of time spent drawing during a session. Indeed, all participants spent a greater proportion of their time drawing in the *free* condition than in the *self-portrait* condition, and we know from a previous study that the durations of the entire drawing sessions were also longer in the *free* condition compared to the *self-portrait* condition (Martinet

et al., 2021). The present study confirms that the addition of instructions limits the proportion of actual drawing time during a session by requiring more reflection time. Regardless of the condition, the proportion of time spent drawing during a session was significantly lower in naive adults. Naive adults, more than any other group, expressed feelings of being judged and apprehension toward doing wrong, and expressed explicitly that they did not know how to draw. Despite these apprehensions and their alleged impacts on performance, the H estimates characterizing the drawing behaviour of naive adults was not different from those of experts. Whether this suggests that different mechanisms can lead to similar fractal patterns in drawing, or that such patterns are not sensitive to subjective experiences during drawing, cannot be determined at this time. However, this does highlight that the differences between young children and adults may not depend strongly on experience or skill but may instead reflect more fundamental ontogenetic processes. Given that the majority of drawings made by 3-year-olds did not exhibit external representativeness, one possible explanation for this difference of complexity with age could be a desire for figuration. Indeed, in adults the process of drawing is intentional and may lead to greater stochasticity in the intermittences between drawing and non-drawing states, due to thought processes and/or tendencies toward representativeness. When an individual produces a figurative drawing, recognizable by an observer, the intentionality of his acts is obvious. However, the fact that only adults fundamentally differ from younger children is noteworthy and can be linked to the theories about the ontogeny of this behaviour. Cohn (Cohn, 2012) speaks of *production scripts* that specify the order in which the elements of the drawing are represented. During the development of drawing skills during childhood, we progressively integrate graphic vocabulary that we would learn to arrange to make

sense by using a particular order of production. We learn to produce figurative schemas by following production scripts that become systematic. Referring to the results we found by applying spatial fractal analysis to the drawing layout (Martinet et al., 2021), we can hypothesize that older children although drawing figuratively, present a more rhythmic and predictable drawing/non drawing intermittence than adults who add many details introducing more variability to the production scripts used in their drawings.

Concerning abstract drawings, non-figuration does not necessarily mean absence of intention. The probable role of drawing instructions can then be evoked. This last consideration leads us to go further in future drawing analyses, asking adults to draw abstract or figurative drawings for comparison with children's' scribbles. As the temporal fractal index is a measure of the temporal complexity of the drawing behavioural sequence, if the two measures of H are different between the abstract drawings made by young children and those made by adults, this would mean that the complexity assessed by the H index could be interpreted as a measure of intentionality more than just figuration. Additional studies would be needed to confirm this.

In the present study, drawing instructions (*free* and *self-portrait*) had no effect on drawing intermittence. Extrapolating from the previous discussion, inviting participants to produce archetypes, more stereotyped drawings of common objects such as a house or a flower, might produce a gradient of temporal patterns depending on the complexity of the task. Perhaps asking subjects to draw a specific object would lead to homogenization of temporal patterns across participants. Asking participants to reproduce a photograph - which would reduce the role of creativity and therefore minimize reflection time but not necessarily simplify the drawing task

- may lead to further variation, and potentially reduced variability between subjects.

The latter would allow measuring the variation in drawing complexity while removing cultural and normative aspects. In this way, it would be possible to reach stronger conclusions about the influence of a directive on drawing behaviour.

Concerning differences in drawing behaviour between individuals, gender can be an influential factor. Previous results have shown such differences in the fields of drawing and writing, particularly with regards to colour utilization, where girls show a more extensive use of colour than boys (Martinet et al., 2021; Turgeon, 2008; Wright & Black, 2013). However, we found no evidence to suggest that drawing intermittence differs between girls/women and boys/men.

Though we observed clear differences in fractal patterns of drawing intermittence between age groups, there remain limitations to the study. For example, difficulty arises from the fact that multiple methods exist to estimate the Hurst exponent, which often leads to conflicting results (Karagiannis et al., 2006). It remains challenging to determine which estimator best suits the type of data being analysed, and this is exacerbated by the fact that estimations themselves vary according to the method used even in simulated time-series with *a priori* seeded Hurst values (Karagiannis et al., 2006). In our study, individuals with high (or low) H according to one method generally exhibited similarly high or low H with other methods, but the variations between groups differed from method to method. Indeed, different methods sometimes produced large differences in H for the same individual (± 0.3). Such large differences interfered with conclusions about whether a drawing sequence was persistent or anti-persistent because in some cases H could be both above and below 0.5, depending on the method.

By applying new methods of analysis, it will be possible to progressively grasp the ontogeny of drawing behaviour. The results of a single study are not sufficient to identify the development of a behaviour as complex as drawing. Only a grouping of clues, each one characterizing one aspect of the behaviour, could make it possible to comprehend the whole. Here we show that it is possible to apply temporal fractal analysis to the behavioural sequences of drawings, allowing to learn more about their structure. Such works seem promising to better understand the ontogeny of drawing behaviour and, by extending this type of analysis to other species, notably great apes, we could learn more about its evolutionary history.

AUTHOR CONTRIBUTIONS

Conceptualization, B.B., L.M., C.S.; Methodology, B.B. and L.M.; Investigation, B.B., L.M. and C.S.; Writing - Original Draft, B.B. and L.M.; Writing - Review & Editing, C.S., A.M, X.M. and M.P.; Funding Acquisition, C.S. and M.P.; Resources, J.H.; Data Curation, B.B. and L.M., Supervision, C.S. and M.P.

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lison Martinet (lison.martinet@iphc.cnrs.fr)

Data and code availability

The dataset generated during this study are available at Mendeley Data: <https://data.mendeley.com/datasets/n672kz24f2/1>

Declaration of Interests

The authors declare no competing interests.

References

- Adi-Japha, E., Levin, I., & Solomon, S. (1998). Emergence of representation in drawing: The relation between kinematic and referential aspects. *Cognitive Development*, 13(1), 25-51.
- Alados, C. L. et al. (1996). Fractal structure of sequential behaviour patterns: An indicator of stress. *Animal Behaviour*, 7.
- Barton, K. (2009). MuMIn: Multi-model inference. [Http://R-Forge. r-Project. Org/Projects/Mumin/](http://R-Forge.r-project.org/Projects/Mumin/).
- Berni, A., Giuliani, A., Tartaglia, F., Tromba, L., Sgueglia, M., Blasi, S., & Russo, G. (2011). Effect of vascular risk factors on increase in carotid and femoral intima-media thickness. Identification of a risk scale. *Atherosclerosis*, 216(1), 109-114.
- Cannon, M. J., Percival, D. B., Caccia, D. C., Raymond, G. M., & Bassingthwaite, J. B. (1997). Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A: Statistical Mechanics and Its Applications*, 241(3-4), 606-626. [https://doi.org/10.1016/S0378-4371\(97\)00252-5](https://doi.org/10.1016/S0378-4371(97)00252-5)
- Cherney, I. D., Seiwert, C. S., Dickey, T. M., & Flichtbeil, J. D. (2006). Children's Drawings: A mirror to their minds. *Educational Psychology*, 26(1), 127-142. <https://doi.org/10.1080/01443410500344167>
- Cohn, N. (2012). Explaining 'I can't draw': Parallels between the structure and development of language and drawing. *Human Development*, 55(4), 167-192.
- Constantine, W., & Percival, D. (2007). *The fractal Package*.
- Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for short time series: A re-assessment of classical methods. *Journal of Mathematical Psychology*, 50(6), 525-544.
- Delignières, D., Torre, K., & Lemoine, L. (2005). Methodological issues in the application of monofractal analyses in psychological and behavioral research. *Nonlinear Dynamics, Psychology, and Life Sciences*, 9(4), 435-461.
- Eke, A., Hermán, P., Bassingthwaite, J., Raymond, G., Percival, D., Cannon, M., Balla, I., & Ikrényi, C. (2000). Physiological time series: Distinguishing fractal noises from motions. *Pflügers Archiv - European Journal of Physiology*, 439(4), 403-415. <https://doi.org/10.1007/s004249900135>
- Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). *Fractal characterization of complexity in temporal physiological signals*. 38.
- Fernandes, D. N., & Chau, T. (2008). Fractal dimensions of pacing and grip force in drawing and handwriting production. *Journal of Biomechanics*, 41(1), 40-46. <https://doi.org/10.1016/j.jbiomech.2007.07.017>
- Freeman, N. H. (1993). *Drawing: Public instruments of representation*.
- Huang, J. (2013). *somebm: Some Brownian motions simulation functions*. URL [http://CRAN. R-project. org/package= somebm](http://CRAN.R-project.org/package=somebm).
- Jolley, R. P., Knox, E. L., & Foster, S. G. (n.d.). *The relationship between children's production and comprehension of realism in drawing*. 26.
- Karagiannis, T., Molle, M., & Faloutsos, M. (2006). *Understanding the Limitations of Estimation Methods for Long-Range Dependence*. 23.
- Longstaff, M. G., & Heath, R. A. (1999). A nonlinear analysis of the temporal characteristics of handwriting. *Human Movement Science*, 18(4), 485-524. [https://doi.org/10.1016/S0167-9457\(99\)00028-7](https://doi.org/10.1016/S0167-9457(99)00028-7)
- Luquet, G.-H. (1927). *Le dessin enfantin. (Bibliothèque de psychologie de l'enfant et de pédagogie.)*.

- MacIntosh. (2014). *The Fractal Primate: Interdisciplinary Science and the Math behind the Monkey*. 25.
- MacIntosh, A. J. J., Pelletier, L., Chiaradia, A., Kato, A., & Ropert-Coudert, Y. (2013). Temporal fractals in seabird foraging behaviour: Diving through the scales of time. *SCIENTIFIC REPORTS*, 10.
- MacIntosh, Alados, C. L., & Huffman, M. A. (2011). Fractal analysis of behaviour in a wild primate: Behavioural complexity in health and disease. *Journal of The Royal Society Interface*, 8(63), 1497-1509. <https://doi.org/10.1098/rsif.2011.0049>
- Mandelbrot, B. B. (1977). Fractals. *Form, Chance and Dimension*.
- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422-437.
- Maria, G. A., Escós, J., & Alados, C. L. (2004). Complexity of behavioural sequences and their relation to stress conditions in chickens (*Gallus gallus domesticus*): A non-invasive technique to evaluate animal welfare. *Applied Animal Behaviour Science*, 86(1-2), 93-104.
- Marmelat, V., Torre, K., & Delignieres, D. (2012). Relative Roughness: An Index for Testing the Suitability of the Monofractal Model. *Frontiers in Physiology*, 3. <https://doi.org/10.3389/fphys.2012.00208>
- Martinet, L., Sueur, C., Hirata, S., Hosselet, J., Matsuzawa, T., & Pelé, M. (2021). New indices to characterize drawing behavior in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Scientific Reports*, 11(1), 3860. <https://doi.org/10.1038/s41598-021-83043-0>
- Mercik, S., Weron, K., Burnecki, K., & Weron, A. (2003). Enigma of self-similarity of fractional Levy stable motions. *Acta Physica Polonica B*, 34(7), 3773.
- Meyer, X., MacIntosh, A. J., Chiaradia, A., Kato, A., Ramirez, F., Sueur, C., & Ropert-Coudert, Y. (2020). Oceanic thermal structure mediates dive sequences in a foraging seabird. *Ecology and Evolution*.
- Meyer, X., MacIntosh, A. J. J., Chiaradia, A., Kato, A., Mattern, T., Sueur, C., & Ropert-Coudert, Y. (2017). Shallow divers, deep waters and the rise of behavioural stochasticity. *Marine Biology*, 164(6), 149. <https://doi.org/10.1007/s00227-017-3177-y>
- Nelson, T. R., West, B. J., & Goldberger, A. L. (1990). The fractal lung: Universal and species-related scaling patterns. *Experientia*, 46(3), 251-254.
- Paraschiv-Ionescu, A., Buchser, E., Rutschmann, B., & Aminian, K. (2008). Nonlinear analysis of human physical activity patterns in health and disease. *Physical Review E*, 77(2), 021913. <https://doi.org/10.1103/PhysRevE.77.021913>
- Peng, C.-K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1), 82-87.
- Peng, C.-K., Mietus, J. E., Liu, Y., Lee, C., Hausdorff, J. M., Stanley, H. E., Goldberger, A. L., & Lipsitz, L. A. (2002). Quantifying fractal dynamics of human respiration: Age and gender effects. *Annals of Biomedical Engineering*, 30(5), 683-692.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2006). nlme: Linear and nonlinear mixed effects models. *R Package Version*, 3(4), 109.
- Rinaldo, A., Rodriguez-Iturbe, I., Rigon, R., Ijjasz-Vasquez, E., & Bras, R. L. (1993). Self-organized fractal river networks. *Physical Review Letters*, 70(6), 822.
- Rutherford, K. M. D., Haskell, M. J., Glasbey, C., Jones, R. B., & Lawrence, A. B. (2003). Detrended fluctuation analysis of behavioural responses to mild acute

- stressors in domestic hens. *Applied Animal Behaviour Science*, 83(2), 125-139. [https://doi.org/10.1016/S0168-1591\(03\)00115-1](https://doi.org/10.1016/S0168-1591(03)00115-1)
- Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F., & Makse, H. A. (2009). Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences*, 106(31), 12640-12645. <https://doi.org/10.1073/pnas.0902667106>
- Seuront, L. (2009). *Fractals and multifractals in ecology and aquatic science*. CRC Press.
- Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J. A., Pitchford, J. W., James, A., Ahmed, M. Z., Brierley, A. S., Hindell, M. A., Morritt, D., Musyl, M. K., Righton, D., Shepard, E. L. C., Wearmouth, V. J., Wilson, R. P., Witt, M. J., & Metcalfe, J. D. (2008). Scaling laws of marine predator search behaviour. *Nature*, 451(7182), 7182. <https://doi.org/10.1038/nature06518>
- Stadnitski, T. (2012). Measuring fractality. *Frontiers in Physiology*, 13.
- Stadnytska, T., Braun, S., & Werner, J. (2010). *Analyzing Fractal Dynamics Employing R*. 29.
- Stroe-Kunold, E., Stadnytska, T., Werner, J., & Braun, S. (2009). *Estimating long-range dependence in time series: An evaluation of estimators implemented in R*. 15.
- Tanaka, M., Tomonaga, M., & Matsuzawa, T. (2003). Finger drawing by infant chimpanzees (*Pan troglodytes*). *Animal Cognition*, 6(4), 245-251. <https://doi.org/10.1007/s10071-003-0198-3>
- Turgeon, S. M. (2008). Sex differences in children's free drawings and their relationship to 2D:4D ratio. *Personality and Individual Differences*, 6.
- Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, 401(6756), 911-914. <https://doi.org/10.1038/44831>
- Weiss, B., Clemens, Z., Bódizs, R., Vágó, Z., & Halász, P. (2009). Spatio-temporal analysis of monofractal and multifractal properties of the human sleep EEG. *Journal of Neuroscience Methods*, 185(1), 116-124. <https://doi.org/10.1016/j.jneumeth.2009.07.027>
- Willats, J. (2005). *Making sense of children's drawings*. Psychology Press.
- Wright, L., & Black, F. (2013). Monochrome Males and Colorful Females: Do Gender and Age Influence the Color and Content of Drawings? *SAGE Open*, 3(4), 215824401350925. <https://doi.org/10.1177/2158244013509254>

II. Article 3 : Fusion d'indices mathématiques pour déchiffrer le dessin

Making drawings speak through mathematical metrics

Publié dans *Human Nature*

Synthèse de l'Article 3

L'étude du dessin en tant que processus et en tant que produit a déjà permis d'en apprendre plus sur le développement du comportement de dessin. Seulement, les indices étudiés ne permettent d'analyser qu'une partie du produit ou du processus qui y est lié. Pour cette raison, cet article propose une méthodologie ayant pour objectif d'analyser des dessins d'humains en regroupant différents indices mathématiques spatiaux et temporels. De ces analyses sont apparues trois catégories de variables. La première dimension, regroupant les métriques spatiales, a été nommée « efficacité ». La seconde dimension, cette fois liée aux couleurs, a été appelée « diversité ». Enfin, la troisième dimension nommée « séquentialité » regroupe les indices temporels. Cette approche méthodologique a permis de mettre en valeur l'intérêt de combiner différents indices liés à différentes dynamiques du dessin.

Making drawings speak through mathematical metrics

Cédric Sueur^{1,2}, Lison Martinet¹, Benjamin Beltzung¹, Marie Pelé³

¹ : Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

² : Institut Universitaire de France, Paris, France

³ : Anthro- Lab, ETHICS EA7446, Lille Catholic University, Lille, France

Corresponding author: Cédric Sueur, cedric.sueur@iphc.cnrs.fr ; IPHC UMR 7178, 23 rue Becquerel, 67087 Strasbourg, France

Abstract:

Figurative drawing is a skill that takes time to learn, and it evolves during different childhood phases that begin with scribbling and end with representational drawing. Between these phases, it is difficult to assess when and how children demonstrate intentions and representativeness in their drawings. The marks produced are increasingly goal-oriented and efficient as the child's skills progress from scribbles to figurative drawings. Pre-figurative activities provide an opportunity to focus on drawing processes. We applied fourteen metrics to two different datasets (N=65 and N=344) to better understand the intentional and representational processes behind drawing, and combined these metrics using principal component analysis (PCA) in different biologically significant dimensions. Three dimensions were identified: efficiency based on spatial metrics, diversity with color metrics, and temporal sequentiality. The metrics at play in each dimension are similar for both datasets, and PCA explains 77% of the variance in both datasets. Gender had no effect, but age influenced all three dimensions differently. These analyses for instance differentiate scribbles by children from those drawn by adults. The three dimensions highlighted by this study provide a better understanding of the emergence of

intentions and representativeness in drawings. We discussed the perspectives of such findings in comparative psychology and evolutionary anthropology.

Keywords: marking gesture, anthropology, evolution, Homo sapiens, comparative psychology

Introduction

Drawing behavior takes time to develop; it proceeds over a series of childhood phases from scribbling in toddlers (Freeman, 1993; Kellogg, 1969) to representational drawing in older children. The act of drawing moves from purely motor pleasure without visual planning (Kellogg, 1969; Luquet, 1927; Golomb, 2003; Thomas & Silk, 1990) to more complex scribbling with aesthetic attentions (Golomb, 2003; Piaget & Inhelder, 2008; Willats, 2005) to full-blown representations. In terms of the development of early mark-making and pre-representational activities (Costall, 1995; Cox, 2005; Matthews, 1984; 1999), three steps have been identified: action representation, romancing, and guided elicitation (Adi-Japha et al., 1998; Matthews, 1984). In humans, action representations, also called gestural drawings, appear both in spontaneous drawing and in response to the request to draw an object (e.g., an airplane). Children may accompany scribbling with verbalizations or sounds such as roaring, which indicates that both their motions and the marks emerging from their drawing instrument simulate the motion of an object. We can then speculate whether children are really scribbling or if they are simply demonstrating an active rather than a figurative mode of representation (Matthews, 1984). The personal actions of children are therefore not random but are intentional and combined with marks and sounds to represent the moving, roaring object (Cox, 2013; Gardner & Wolf, 1987; Wolf, 1988; Wolf & Perry, 1988). So the drawing has representativeness for the drawer (i.e., internal representativeness) even if this is not the case for an external observer (i.e., external representativeness (Martinet et al., 2021)). Abstract art made by artists also has internal representativeness but sometimes not external because the art is not figurative. “Romancing” refers to instances in which children name a scribble with an object but an observer has difficulty finding a graphic

resemblance between the scribble and the object the child claims to have drawn. Naming takes place either spontaneously or when elicited by an adult and can occur before or during the drawing, or after its completion. Action representations and romancing cannot be observed in children before they are two to three years old, or in children with certain psychopathologies such as autism (Charman & Baron-Cohen, 1993; Jolley et al., 2013). In these two situations, it is therefore difficult to establish if a drawing is goal-directed, having a meaning and an intentional representation. Evidence of intention is provided during the third step, called guided elicitation, when children show no representational intention in their free drawings but produce figurative drawings when assisted. However, even if it is not always easy to demonstrate intentional prerepresentational activities, we can predict without difficulty that the marks produced are increasingly goal-oriented and efficient as the child progresses from scribbles to figurative drawings. Rather than studying the finished drawings, we should focus on the presence of pre-representative activities on drawing processes (Costall, 1995; Cox, 2005; Martinet & Pelé, 2021; Matthews, 1999). The aim of this paper is to demonstrate that mathematical metrics can give cues about intention and representativeness. Could these metrics enable us to measure the pre-representative activities that occur in the early stages of drawing? Different methods have been used to answer this question of representativeness and goal directedness beyond drawing (Desmet et al., 2021; Urban, 2004). Published topics range from the kinematic aspects of scribbling to the precursors of graphic representation, including authors who compare curved lines in drawings to mathematical laws.

Children tend to attribute representational meanings (e.g., an airplane) to angular curves and nonrepresentational meanings (e.g., a line) to smooth curves that they

had just drawn (Kellogg, 1969). However, these studies are not objective given that the authors asked the children what their drawings represented a posteriori. Moreover, this methodology cannot be applied to nonverbal subjects (whether human or nonhuman primates) who are unable to explain their drawings. Thus, only one study to date has used methodology permitting the comparison of drawing abilities in humans and nonhuman primates or made it possible to understand the evolution of drawing in children and in other primates (Martinet et al., 2021). This paper describes the use of different mathematical metrics with the aim of objectively and quantitatively measuring intention and representativeness in drawing. Just because the observer cannot identify an object or an intention in the drawing, since it is made by a toddler or it is abstract art, does not mean that intention and representativeness are absent. Is there a mathematical way to identify these concepts? Recently developed techniques make it possible to consider ethology as a physical science and apply quantitative measures in this discipline (Brown and De Bivort, 2018). Although simple drawing measures such as the number of colors can be used (Zeller, 2007), they provide few cues about the intention behind the drawing. We aim to go further by using mathematical measures to fulfil this goal. Metrics are already used to understand whether movements are optimal or evaluate the extent to which behavioral sequences are complex and predictable in animals, including humans. For instance, Martinet et al. (2021) and Beltzung et al. (2022) used spatial and temporal fractal analyses, which had previously been applied to understand optimal movements and optimal behavioral sequences of animals searching for food (Bartumeus et al., 2002; Meyer et al., 2017; 2020; Reynolds, 2008), and found an increase in complexity and efficiency in humans compared with chimpanzees, but also an increase with age in humans. Other metrics such as entropy

(Ebeling et al., 2002; Kershenbaum, 2014) or the Gini index (Debache et al., 2019; Planckaert et al., 2019) were also used to understand the distribution or complexity of different behaviors (e.g., activity, food exchange, vocalization) from ants to humans. We used a total of fourteen metrics to enrich our understanding of the intentional and representational processes behind drawing. These metrics are detailed in Table 1 along with definitions and predictions. This paper seeks to combine all of these metrics in different dimensions using principal component analysis. PCA is used to extract and visualize important information contained in a multivariate data table by combining metrics to form a biologically or psychologically significant dimension, as has already been shown for personality (Bousquet et al., 2015; Wolf and Weissing, 2012) or sociality (Viblanc et al., 2016). Here, we expect metrics to be combined and form dimensions that correspond to representativeness (at least internal, meaning from the point of view of the drawer, but not for the observer) and show evidence of anticipation (predictiveness of the sequence of behaviors), performance, or aestheticism in the drawing (Dissanayake, 2001; Matthews, 2003; Watanabe, 2012; Wolf and Perry, 1988). Temporal metrics and some spatial metrics based on fractal theory could indicate representativeness and anticipation while the use of colors and space could indicate a sense of aestheticism. The two objectives of this paper are to assess (1) whether mathematical measures can reveal patterns in drawing suggesting intention of representativeness and (2) whether these patterns and measures are linked to the age of the participants. More specifically, we hypothesize that objective and quantitative patterns in drawings will provide cues about the intentions and representativeness of the drawer, even if the observer fails to perceive an object or entity in the drawing. We used two datasets of drawings with different instructions in order to generalize results. We asked adults

to make drawings of different patterns, including an abstract one. Then, we asked children and adults to draw with different instructions (e.g., self-portrait, free drawing, scribbling) in order to reveal these dimensions. We expect that drawings as scribbles would not exhibit similar dimensions if they are made by adults or toddlers, allowing us to attribute concepts behind each dimension. More precise predictions are given below for each metric. We expect children's drawings to become more complex with age, being more efficient in terms of representativeness, including more color diversity and more sequences.

Table 1 Metrics used to understand drawing complexity with their definition and their meaning in the context of drawing

Type	Metric	Definition	Meaning (“it measures”) or expectation (“we expect”)	References
Spatial metrics	μ MLE	Spatial fractal metric, maximum estimate power coefficient of the drawing length distribution	It measures drawing efficiency, from random trajectories to optimal trajectories indicating a representativeness	Edwards et al., 2007; Sueur, 2011; Sueur et al., 2011; Viswanathan et al., 2008
	Drawing distance	Total distance of drawing, from the first point to the last, in pixels	We expect long distance drawings to be more representative or contain more details than short distance drawings. However, it can also mean deterministic drawing, (i.e. no intention to represent anything).	Janmaat et al., 2021; Mitani and Nishida, 1993; Papi et al., 1995
	Angle distribution metric	PCA dimension based on the coefficients of the cubic survival function of angle distribution (from 0° to 180°). Close to the Gini index in its calculation.	We expect homogeneous distributions of angles (low values of this metric) to indicate randomness (scribbles) whilst heterogeneous distributions should be linked to goal directedness (i.e., representativeness)	Bartumeus et al., 2008; Benhamou, 2004; Gurarie et al., 2016; Potts et al., 2018
	Minimum convex polygon	Minimum polygon covering all drawing points and giving the percentage of drawing cover on the screen	We expect high covering to inform about representativeness but also the play /emotional interest in drawing	Gaston and Fuller, 2009; Nilsen et al., 2008

Table 1 (continued)

Type	Metric	Definition	Meaning (“it measures”) or expectation (“we expect”)	References
Temporal metrics	Hurst index	Temporal fractal metric, measure of the long-term process in temporal sequence	It measures the temporal complexity of drawings sequences, from deterministic to complex, including both sequences of drawing and of non-drawing.	MacIntosh, 2014; MacIntosh et al., 2011
	Temporal Gini index	Measure of the inequality among values of temporal drawing sequences distributions	We expect high Gini index, meaning unequal distribution of sequences to give an idea about intention and anticipation in drawing	Debache et al., 2019; Planckaert et al., 2019
	Entropy index	Measure of the temporal uncertainty of drawing	It measures the stochasticity or predictability of the drawing state. We expect a high entropy index to be linked to more representative drawing, including anticipation	Ebeling et al., 2002; Kershenbaum 2014; Leff, 2007
	Drawing test time	Total drawing time (including drawing time and non-drawing time)	We expect a long duration to inform about thinking about drawing, i.e., intention and representativeness	
	Number of sequences	Number of drawing and non-drawing sequences during the test	We expect a high number of sequences to give an idea about goal-oriented behaviors, meaning intention and representativeness	
	Drawing speed	Speed of drawing, which is the drawing distance in pixels divided by the time of drawing	Speed is used as a measure of goal directedness or knowledge (i.e., in this context, mastering)	Byrne et al., 2009; Noser and Byrne, 2014; Suceur, 2011
	Draw- ing time proportion	Drawing time divided by test time	We expect a high drawing time proportion to inform on thinking about drawing, i.e., intention and representativeness	
Color metrics	Mean colorimetric profile	Mean distribution of intensity levels for the Red, Green, or Blue colors, respectively, and after removing the white (screen) color on the parts covered by drawing	It measures the mean spectrum of colors used, from dark to light.	
	Standard deviation of the colorimetric profile	Standard deviation of the distribution of intensity levels for the Red, Green, or Blue colors, respectively, on the parts covered by drawing	It measures the diversity of the spectrum of colors used. It is different from the number of colors as it takes also how much these colors are different.	
	Number of colors	Number of colors used from the ten proposed colors.	We expect the number of used colors to give an idea about the aestheticism but also inform about the play interest in drawing.	Martinet et al., 2021

Material and Methods

Dataset

All methods were performed in accordance with the relevant guidelines and regulations in France, checked and approved by the research ethics committee of Strasbourg University (Unistra/CER/2019-11). We used two datasets. The first is composed of drawing conditions expected to yield the greatest variation in drawings characteristics and then affecting the measures we used. The second is composed of drawing conditions used to test the effect of age and gender. The datasets are as follows:

1. We asked 13 adults (6 men, 7 women) from 21 to 29 years of age to make five drawings. These adults were considered “naive” insofar as they had never taken drawing lessons and did not draw as a hobby. These participants were students of the research institute where the authors worked. Each drawing corresponded to an instruction that was specifically designed to produce a range of drawings: (a) draw something with scribbles; (b) draw something with circles; (c) draw something with different angles; (d) draw something with a starry sky; and (e) draw something with fan patterns (fan patterns are defined as straight lines with sharp angles due to repetitive actions of hand, usually from right to left). The reasons for the choice of these five instructions are detailed with the list of metrics below. Examples for each instruction are given in Fig. 1. This dataset was collected in 2020 and is composed of 65 drawings.

2. The second dataset included both children and adults. A group of 144 children from three to ten years of age was split into 18-20 children per one-year age interval. Boys and girls were equally represented in these categories except in the youngest category, which was composed of 5 girls and 15 boys. In this study, the aim of

including children was not to test for differences by age (that is the next step) but to have a large panel of different drawings. The adult group was composed of 41 adults (21 men, 20 women) 21 to 60 years of age who were either naive or expert drawers. The latter were art school students and professional illustrators. Participation was voluntary for adults and subject to parental consent for children. None of the adults exhibited severe physical or cognitive traits that could impact drawing capacities. These two conditions were visually and verbally checked with participants. According to Martinet et al. (2021), all participants were asked to draw in two different conditions: free drawing (*draw what you want*: the experimenter told the subject that they could draw whatever they wanted, with no further instructions) and self-portrait conditions (*draw yourself*: the experimenter instructed the subject to draw themselves). The dataset was collected in 2018 and 2019. Further information about this dataset (i.e., methodology, examples of drawings, and video footage of hand movements) is given in Martinet et al. (2021). A total of 370 drawings was initially collected for this dataset; however, some data were lost during the recording processes. The final dataset therefore contains 344 drawings, but the distribution per age or sex categories was not affected. Three of the naive subjects are present in both datasets.

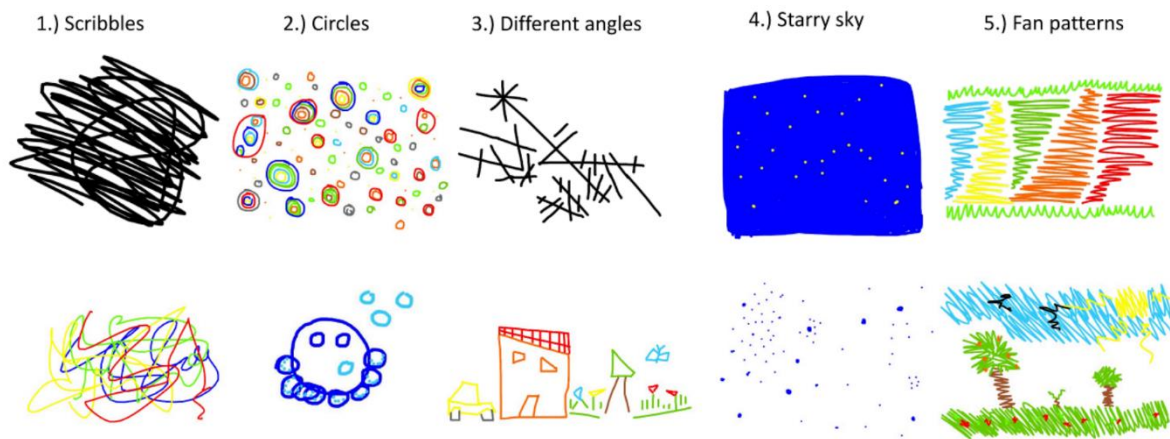


Figure 1. Examples of the five instructions we gave to participants for dataset 1. 1.) Make a drawing with scribbles; 2.) Make a drawing with circles; 3.) Make a drawing with different angles; 4.) Make a drawing with a starry sky; 5.) Make a drawing with fan patterns.

Experimental design

The experimental design is similar to the one described in Martinet et al. (2021).

Habituation phase : Each participant was given a touchscreen tablet (iPad Pro, 13-inch, version 11.2.2, capacitive screen reacting to the conductive touch of human fingers) and was told to draw on it so they could be familiar with how it worked and how to change the color(s) they wanted to use. Drawing with fingers enabled very young children who had not yet mastered the use of a pencil to be included. A panel consisting of ten different colors was displayed at the bottom of the screen, and the participant could select a color by clicking on it. When they clicked on a different color in the panel, any subsequent drawing was produced in that color. Children were habituated to the touchscreens the day before the tests to avoid overstimulation. Adults were tested immediately after familiarizing themselves with the tablet.

Testing phase : Each child was individually tested during school hours at school, either in their classroom (for the 3-year-olds) or in the staff room (for the older children). The experimenter (LM or MP) stayed during the test but kept at a distance to avoid influencing the children. Adults were tested individually in a room at the

research institute (for naive participants) or at the art school (for expert drawers). Adult participants were left alone in the room. A camera recorded the hand movements of all participants while drawing in case we needed to check for any problem during the session (interruption of the drawing, involuntary tracings, etc.). No time limit was applied.

Data analysis

For each drawing, the spatial coordinates X and Y of every point of the lines drawn as well as their time coordinates [min; s; ms] and the color used were automatically scored on a datasheet (text file) by the iPad Pro. These data enable us to calculate spatial, temporal, and color metrics for each drawing (Table 1). Details of metrics, their calculation, and the range of values for each instruction for dataset 1 can be found in the ESM. The number of sequences is correlated to the number of lines in the drawing. We retain the number of sequences as data so the temporal sequences can be analyzed using the Hurst index and analyzed in parallel with the duration of each sequence. For the first dataset, we expect different values based on the instructions provided to the participants:

1. Scribbles are usually more or less random patterns in toddlers, with no figurative intention but only locomotor pleasure. However, we expect adults to be less spontaneous in their scribbles. Scribbles should be done with few posing (nondrawing sequences) and drawing time should be short because there should be not figurative purpose or goal in this drawing condition in adults. For this reason, the scribbles are not expected to show sharp angles and straight lines, so we predict the observation of a small μ MLE (Table 1) and a small angle distribution metric, but also a large minimum convex polygon. The drawing session duration and the number of sequences

should be low but the drawing speed should be high. Finally, scribbles should have few colors.

2. We imagined drawing with circles would not result in sharp angles (but rather obtuse ones) nor straight lines, so we expect to obtain a small μ MLE and a small angle distribution metric. We have no presupposition for the minimum convex polygon. Likewise, we cannot predict the drawing session duration or the number of sequences, whether in terms of speed or the number of colors used.

3. We imagined drawings with different angles to obtain large angle distributions as well as large distributions of line lengths, meaning an intermediate μ MLE. We have no hypothesis for the minimum convex polygon. The number of sequences should be high, corresponding to the different angles/lines, but we cannot predict the duration of the drawing session, the drawing speed, or the number of colors used.

4. Adults drawing a starry sky should draw randomly small stars on the paper sheet, separated by long distances. We therefore expect drawings of a starry sky to contain different angles, but mostly sharp ones depending on the stars' shape. Long distances between stars but use of short lines to draw the stars should yield a high μ MLE. The minimum convex polygon, the number of sequences, and the drawing session duration should all be high since adults should draw the sky on the entire paper sheet. The number of colors should be low and the colors should be light unless the participants drew a dark sky.

5. Fan patterns are parallel broad zigzag strokes, meaning stereotyped or repeated sequences of straight and long lines separated by sharp angles (Pelé et al., 2021). We have no prediction for the minimum convex polygon. Speed should be high given the findings in previous studies on fan patterns (Kellogg, 1969; Zeller, 2007). The

number of sequences should be high in relation to the different angles/ lines. However, we have no prediction for the duration of the drawing session or for the number of colors used.

Statistical analysis

As preliminary results, we analyzed whether and how each metric differs between drawings for each instruction. This was achieved using ANOVA, or a Kruskal-Wallis test when ANOVA conditions could not be met (i.e., non-Gaussian distribution). Pairwise comparisons were realised when ANOVA or Kruskal-Wallis tests were significant (the “TukeyHSD” function of the R base package and the “kruskalmc” function of the “pgirmess” package (Giraudoux et al., 2018), respectively). Only differences with $p < 0.05$ were reported.

Analyses were carried out in three main steps using correlation analyses and principal component analyses: Analysis of the first dataset (as described below), analysis of the second dataset, following the same procedure as in step 1, and a final combined analysis of the datasets to generalize our results.

The first step, on dataset 1: a correlation analysis was carried out with the R package “PerformanceAnalytics” (Carl et al., 2010; Peterson et al., 2018) on all metrics to identify those that were highly correlated ($R > 0.9$). Following this correlation analysis, we removed the drawing duration proportion metric, which was highly correlated to the Gini index. Most of the variables were also influenced by the drawing test duration metric. We therefore decided to correct all the variables by carrying out a linear regression, using each metric as a response variable and the drawing test time as a factor. We took the residuals from this linear regression, which corresponds to any variance of each point that was not explained by the drawing test

duration. A Principal Component Analysis (Budaev, 2010; Holland, 2008) with Varimax rotation was then carried out using the R package 'Psych' (Revelle, 2011; Revelle & Revelle, 2015). Variables are automatically corrected to be comparable (mean and range). Three dimensions were set up. Varimax rotation is used to simplify the expression of a particular subspace in terms of just a few major items each. This means that the Varimax rotation applies the variables to each dimension in turn in order to maximize the explained variance. We examined the loadings of each variable on each dimension. The loadings are interpreted as the coefficients of the linear combination of the initial variables from which the principal components are constructed. The loadings are equal to the coordinates of the variables divided by the square root of the eigenvalue associated with the component. We removed variables for which loadings are less than 0.4, which indicates a weak contribution to each dimension and to the total explained variance. After this removal, we renewed PCA with Varimax rotation and analyzed the results.

The second step, on dataset 2: We followed the procedure described for dataset 1.

The third step, on dataset 1 and 2: We compared the variables contributing to each of the three dimensions for the two datasets. We removed the variables that did not contribute to the same dimensions between dataset 1 and dataset 2 and performed a PCA with Varimax rotation on both datasets, checking that we obtained similar or better results and according to the parsimony rules (see Results). We did this as a way to compare all drawings using the same method and to generalize this analysis. These results were then compared to assess whether our procedure might be generalized to any dataset. PCA dimensions were compared via a Pearson correlation test. Finally, the same PCA procedure was used to combine both datasets and compare the scribbles made by adults following the instruction we gave ("draw

something with scribbles”) and the “natural” scribbles of 3-year-old children. A Mann-Whitney test was performed to compare both categories in each dimension.

Finally, we tested the effect of age and gender in dataset 2. We ran generalized linear models (GLMs) with each dimension as the dependent variable and the gender of participants, age categories, and conditions (free drawing or self-portrait) as independent variables. We also added the interactions of gender-condition and group- condition as independent variables to the model. Normality and homoscedasticity of residuals were verified graphically. We then performed pairwise comparisons with Benjamini-Hochberg (Verhoeven et al., 2005) correction for significant independent variables. All analyses were performed in RStudio 1.4.1103 (Allaire, 2012; Racine, 2012). The significance threshold was set to $\alpha=0.05$.

Results

Preliminary results on dataset 1: the details of tests and pairwise comparisons between instructions (dataset 1) for each metric are available in the ESM. Three metrics showed similar values for the five instructions given to participants: angle distribution metrics, the mean colorimetric profile, and the standard deviation of the colorimetric profile. The minimum convex polygon was smaller in drawings composed of different angles than in fan pattern drawings. The number of colors was higher for the fan pattern instruction than in the different angles drawing. The scribble instruction results are different from all others in terms of drawing test time (i.e., lower), entropy (i.e., lower), number of sequences (i.e., lower), the Gini index (i.e., lower for all instructions except the starry sky), and the Hurst index (i.e., higher except in comparison with the fan pattern instruction). The starry sky instruction is linked to a longer drawing distance than all other instructions except

fan patterns, while the different angles instruction is linked to a shorter drawing distance. The different angles instruction has a lower drawing time proportion and a higher Gini index than all other instructions except “draw circles.”

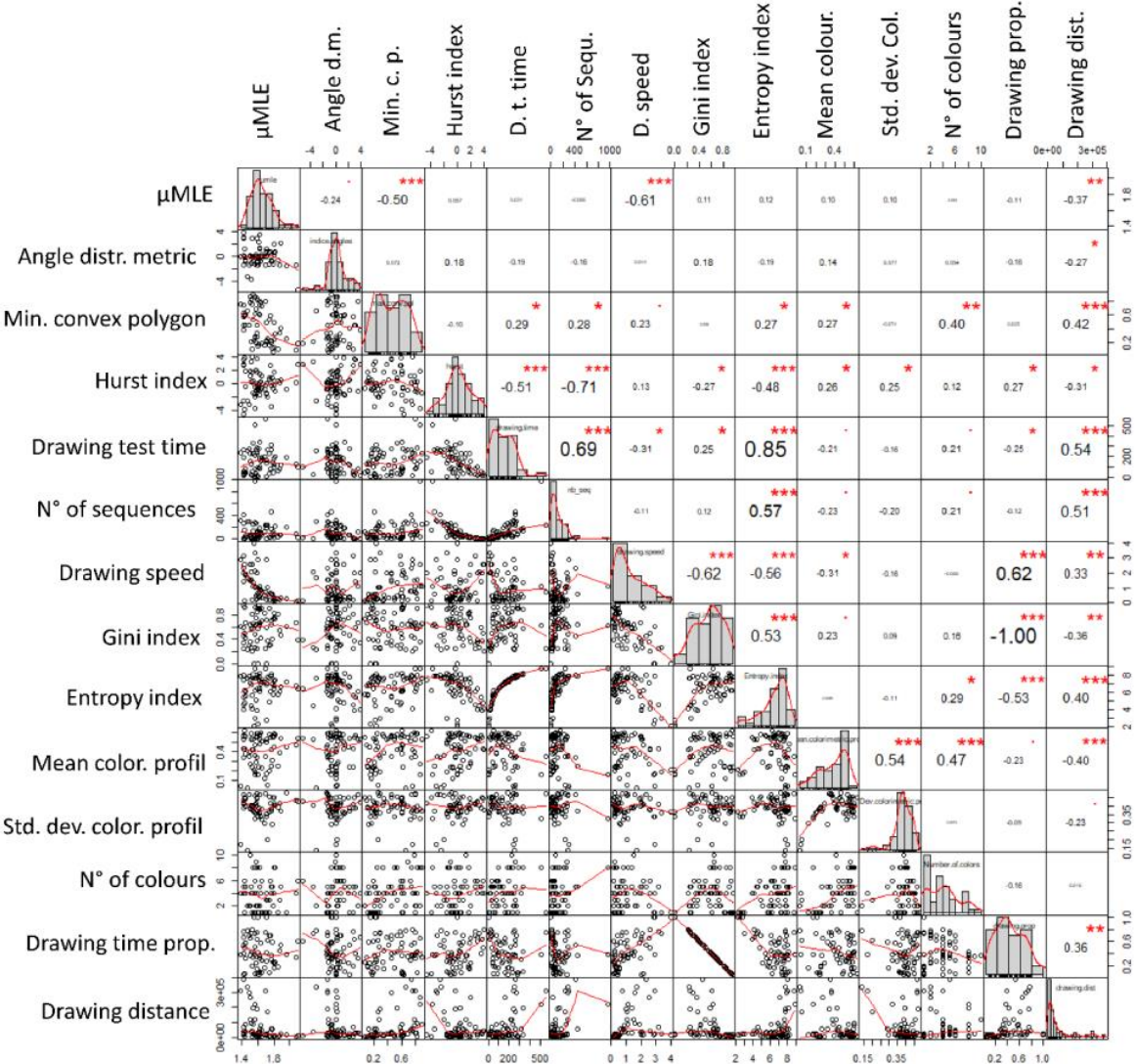


Figure 2. Correlation chart of the 14 metrics for dataset 1. The diagonal of the graph provides the distribution of each metric, whilst the bottom left and the top-right provide the correlation figure and the correlation coefficient between two metrics, respectively. Statistical value is given with the correlation coefficient: * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

First step on dataset 1: The results for the correlation analyses of metrics for the first dataset are shown in Fig. 2. Drawing duration proportion is highly correlated ($r = -1$) with the Gini index. We decided to remove drawing duration proportion as a variable. Moreover, as we could expect, 10 of the 13 variables are correlated with

drawing session duration. The latter is the most correlated with other variables. We then corrected all remaining variables according to the drawing duration. The correlation chart describing these corrected metrics is shown in Figure S15. This step was followed by a PCA with Varimax rotation. The total explained variance is 55.7% (dimension 1=24.5%, dimension 2=17.2%, dimension 3=14%). Three variables have a loading below 0.4 for all three dimensions (details in the ESM, Table S1), namely the angle distribution metric, the drawing session duration and the standard deviation of the colorimetric profile. We therefore removed these three variables from the dataset and carried out another Varimax rotation PCA. The total explained variance of this new PCA is 69.8% (dimension 1=31.9%, dimension 2=20.7, dimension 3=17.2%). Each metric shows a higher loading value in one dimension, unlike the two others (Table 2). We can thus attribute each metric to one dimension as follows: dimension 1 (μ MLE, drawing speed, Gini metric, entropy metric, drawing distance); dimension 2 (minimum convex polygon, mean colorimetric profile, number of colors); dimension 3 (Hurst index, number of sequences). Examples of dataset 1 drawings scaled to the three dimensions are given in Figure S16a-c.

Second step on dataset 2: We followed the same steps as those described for dataset 1. Results for the correlation analyses of metrics of the second dataset are shown in Figure S17. The results of dataset 2 are comparable to those of dataset 1: the drawing duration proportion was highly correlated to the Gini index ($R=1$), and was therefore removed to correct other variables by the drawing duration. The correlation chart depicting these corrected metrics is shown in Figure S18. This step was followed by a Varimax rotation PCA. The total explained variance is 55.3% (dimension 1=18.7%, dimension 2=18.4%, dimension 3=18.2%). As in dataset 1, the angle distribution and the drawing session duration have a loading below 0.4 for each dimension. However,

the standard deviation of the colorimetric profile has a loading equal to 0.87 for dimension 1. We removed this variable to ensure a fit with the results of dataset 1; this does not change the variance explained (64.6% with versus 64.7% without) or the contributions of other metrics to the different dimensions. We carried out another Varimax rotation PCA. The total explained variance of this new PCA is 64.7% (dimension 1=24.2%, dimension 2=23.5%, dimension 3=17.1%). Each metric shows a loading value higher in one dimension, unlike the two others (Table 2). We can thus attribute each metric to one dimension as follows: dimension 1 (μ MLE, drawing speed, drawing distance, minimum convex polygon), dimension 2 (mean colorimetric profile, number of colors), dimension 3 (Hurst index, number of sequences, Gini metric, Entropy metric).

Table 2. Loadings of the metrics (after loadings selection) on the three Varimax rotation PCA dimensions of dataset 1 and dataset 2. Bold values indicate the dimension in which the metric is retained in each dataset. Grey highlights indicates similar results for both datasets. Axes of Dimension 1 are inversed between dataset 1 and dataset 2, but results and correlations are similar.

	Dataset 1			Dataset 2		
	Dim. 1	Dim. 2	Dim. 3	Dim. 1	Dim. 2	Dim. 3
μ MLE	0.608	-0.389	-0.303	-0.781		-0.276
min. conv. pol	-0.374	0.783	0.134	0.666	0.404	-0.133
Hurst index		0.214	-0.881	-0.138		-0.911
No. of sequences	-0.131		0.79			0.775
Drawing speed	-0.914		0.104	0.867	-0.111	0.245
Gini index	0.784	0.19	0.211	-0.14	0.42	0.685
Entropy index	0.659	0.363	0.212			0.45
Mean color. profile	0.437	0.711	-0.191	-0.168	0.857	
No. of colors	0.11	0.759	-0.144	0.181	0.607	
Drawing distance	-0.754		0.223	0.708	-0.488	-0.296

Table 3. PCA loadings of the metrics similar to datasets 1 and dataset 2. **Bold values indicate dimensions in which the metric is retained in each dataset**

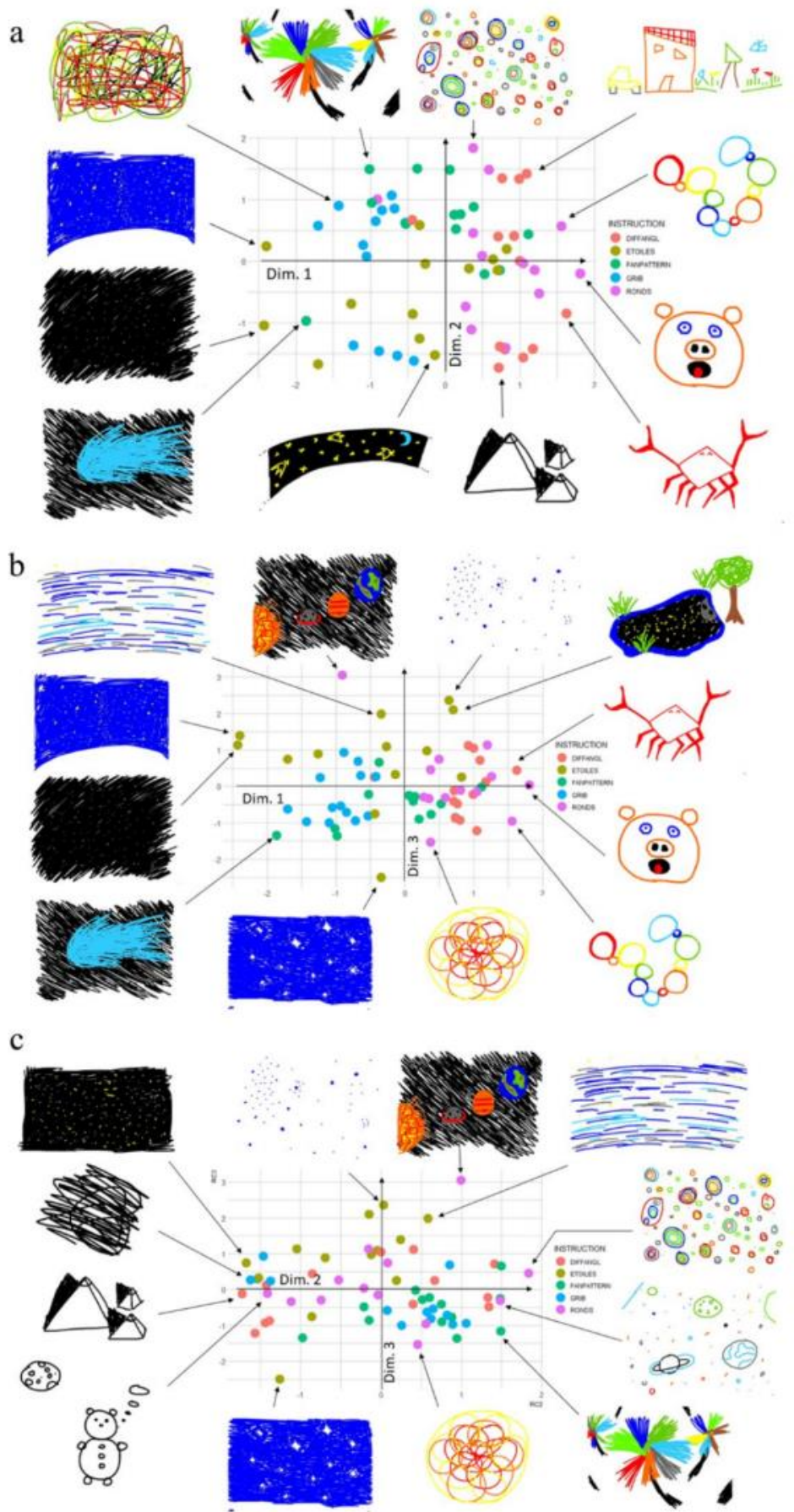
	Dataset 1			Dataset 2			Datasets 1 and 2		
	Dim. 1	Dim. 2	Dim. 3	Dim. 1	Dim. 2	Dim. 3	Dim. 1	Dim. 2	Dim. 3
μ MLE	0.812			0.814	-0.117	0.246	0.778	-0.14	0.252
Hurst index		0.248	0.872	-0.143		0.902			0.905
No. of sequences	-0.188	0.112	-0.886			-0.898			-0.896
Drawing speed	-0.894	-0.133		-0.9		-0.216	-0.88		-0.136
Mean colors profile	0.317	0.81		0.291	0.777		0.315	0.777	
No. of colors		0.909		-0.178	0.795		-0.141	0.824	
Drawing distance	-0.793	-0.194		-0.746	-0.371	0.296	-0.775	-0.312	0.21

Third step on datasets 1 and 2: Seven of the ten retained variables belong to the same dimension in the PCAs carried out for datasets 1 and 2 (Table 2), and have quite similar loadings. However, three variables (minimum convex polygon, entropy index and Gini index) are not found in the same dimensions in the two datasets. When these three variables were removed from the PCA, we obtained similar results with comparable loadings per metric (Table 3) and 77.5% of the variance was explained for dataset 1 (dimension 1=31.9%, dimension 2=23.2%, dimension 3=22.4%), whilst 77% of the variance was explained for dataset 2 (dimension 1=31%, dimension 2=26.1%, dimension 3=19.9%). Reducing the selection of variables from ten to seven does not substantially change the classification of drawings, as the values of the three PCA dimensions are highly correlated between the first and the third step in dataset 1 (RC1: $t=18.942$, $df=63$, $p<0.0001$, $r=0.92$; RC2 : $t=14.357$, $df=63$, $p<0.0001$, $r=0.87$; RC3 : $t=-15.764$, $df=63$, $p<0.0001$, $r=0.89$). When we combined both datasets, we obtained similar results to those obtained in separate analyses of dataset 1 and dataset 2, with 77.5% of the variance explained (dimension 1=30.2%, dimension 2=25%, dimension 3=20.2%; Table 3). Examples of dataset 1 drawings scaled on the three dimensions are given in Fig. 3. Finally, we compared the three

dimensions between scribbles of dataset 1 (made by adults) and scribbles of dataset 2 (made by 3-year-old children only, as scribbles become rare from the age of four onwards) in order to show that the PCA is able to show differences between drawings unable to see with our human eyes. Mann-Whitney tests showed that dimension 1 differs between datasets 1 and 2 ($w=29$, $p=0.0066$) whilst there is no significant difference between the two sets for dimension 2 ($w=107$, $p=0.122$) and dimension 3 ($w=105$, $p=0.152$) (Fig. 4). Moreover, Fig. 4 shows that data are more dispersed for the three dimensions in adults' scribbles compared to children ones. Mann-Whitney test for each metrics in each dimension are detailed in the supplementary material (Table S2). Effect of age and gender on dataset 2: With respect to the GLMs, the model selection of each dimension and the results for each variable are detailed in the ESM (Tables S3-S8). Interactions were only present in the models which best explained dimension 2 variance, but were not significant (conditions: gender, $p=0.139$). Gender did not influence any of the three dimensions ($p>0.097$). The drawing condition only influenced the dimension 2 ($p=0.016$), with higher values for the free drawing indicating a higher number and spectrum of colors were used. Age categories influenced all three of the dimensions (Fig. 5). Pairwise comparisons (Table S9) showed that dimension 1 values of 3-year-old children were lower than that of 5-, 7-, and 9-year-old children ($p<0.044$), as was the ones of expert adults ($p<0.03$). This means that we observed an increase and then a decrease with maximum efficiency around 9 years of age. Dimension 2 values (Table S10) were higher in 3-year-old children than in other categories ($p<0.03$), except that of 4-year-old children and novice adults. A higher dimension 2 was noted in 4-year-old children compared with that of 7-, 9-, and 10-year-old children and expert adults ($p<0.004$). A lower dimension 2 was noted in 9-year-old children compared with that of 3-, 4-,

5-, and 8-year-old children and novice adults ($p < 0.03$). Novice adults showed a higher dimension 2 than 9-year-old children and expert adults ($p < 0.03$). There was a general decrease and then increase in the dimension 2 as young children, older children, and novice adults added more colors than middle-aged children and expert adults. Finally, dimension 3 values (Table S11) of novice adults were higher than that of all other age categories ($p < 0.004$). Expert adults showed higher dimension 3 values compared with 3-, 4-, and 7-year-old children ($p < 0.038$).

Figure 3. Examples of dataset 1 (third step) drawings (a) according to dimension 1 and dimension 2, as provided by the PCA. Dimension 1 may represent representativeness and dimension 2 may represent diversity. (b) according to dimension 1 and dimension 3, as provided by the PCA. Dimension 1 may represent representativeness and dimension 3 may represent periodicity. (c) according to dimension 2 and dimension 3, as provided by the PCA. Dimension 2 may represent diversity and dimension 3 may represent periodicity



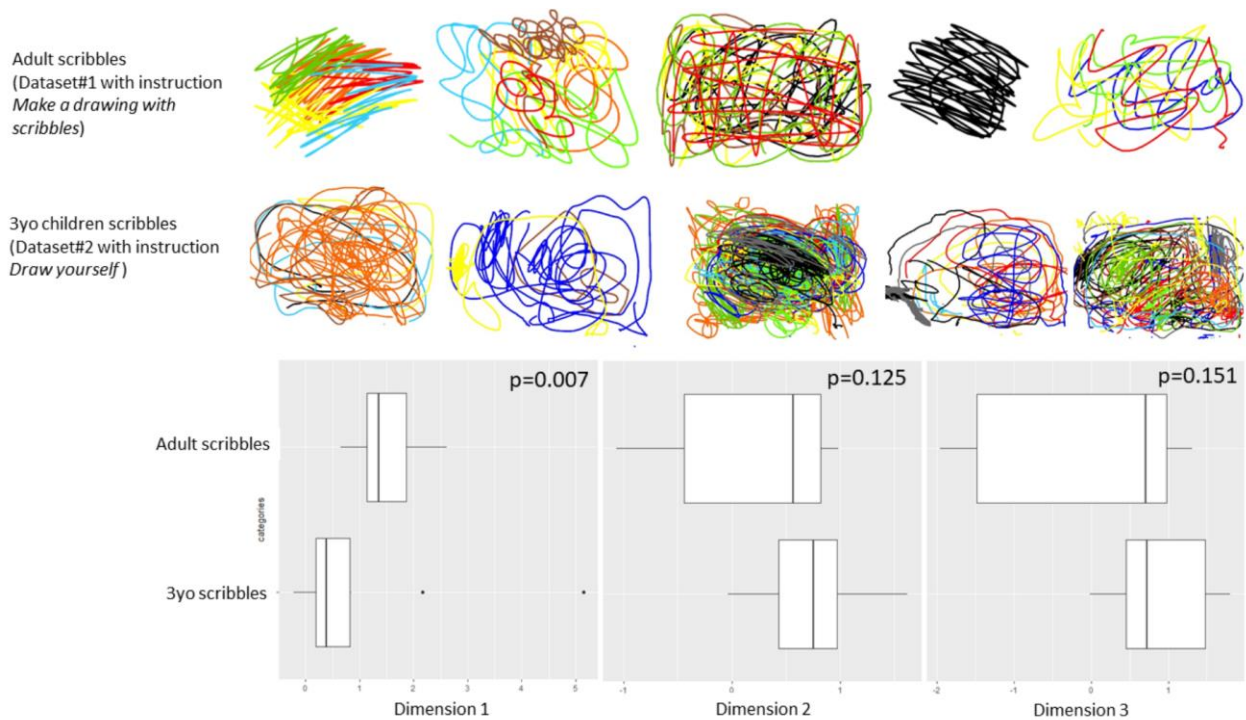


Figure 4. Examples of scribbles for dataset 1 (adults) and dataset 2 (3-year-old children) and boxplot of dimensions 1 to 3 for these two categories.

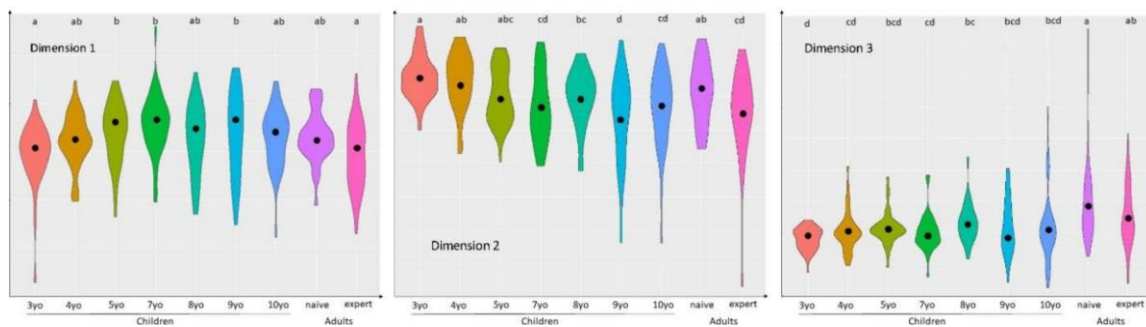


Figure 5. Violin plots representing dimension 1, dimension 2, and dimension 3 across age categories of dataset 2. Violin plots represent the distribution of data, and the black dots represent the mean. Age groups with the same letter belong to the same group according to pairwise comparison tests (see details in ESM).

Discussion

We used several mathematical metrics to characterize drawings and assess whether they can give cues about representativeness and intention. Principal component analyses helped us to organize these metrics in three dimensions in a first dataset. The analyses on this first dataset were then confirmed with the analyses carried out on a second dataset, thus allowing us to generalize our method of characterizing

drawings and the subsequent results. This study is an important step in the analysis of drawings as it is the first time that such a high number of mathematical indices are used to analyze the cognitive processes behind creativity. This discussion seeks to understand which process corresponds to each dimension provided by the PCA.

The choice of two datasets that each involves different drawing instructions proves to be the right protocol to obtain variations in each metric. The colors used to draw are influenced by the instructions we gave. The choice of these instructions was designed to produce a variety of shapes and lines influencing spatiotemporal metrics. However, variation in used colors is still high and two color metrics (i.e., number of colors and mean colorimetric profile) partly explains variance in different dimensions. Nevertheless, the standard deviation of the colorimetric profile does not provide any information about the drawing, as the choice of colors variable is more a question of personal preference than characteristic of any representative process. Similarly, and contrary to what we expected, the angle distribution metric does not differ between the instructions of dataset 1 and did not play any role in explaining variance in the PCAs of dataset 1 and dataset 2. This result might be due to the issuing of an unsuitable instruction, thus leading to false negatives. However, as we obtained similar results between dataset 1 and dataset 2, the explanation should be more in the drawing process itself where, whatever the objects, their representation produces similar angle distributions. The study of turning angles is often used when assessing the optimality of animal movements, but it is limited to differentiating random movements from goal-oriented and directed ones (Reynolds, 2008; Sueur, 2011; Sueur et al., 2011). The scale used to analyze angles in our study is possibly too limited to obtain significant results. After correcting all variables according to the drawing session duration, no difference is seen in the latter between

instructions, nor did it play a role in explaining PCA variance in either of the datasets. This means that representativeness and aestheticism in drawing are not directly linked to drawing duration but more to all the other processes (e.g., the drawing traits length and the number of sequences depending on the number of objects) that influence the duration of drawing. In dataset 1, scribbles were the only instruction leading solely to nonrepresentative drawings, despite the fact that the definition of a scribble can be unclear. Other instructions mainly resulted in the drawing of objects or animals. It is difficult to assess whether it is the instruction itself or the process of drawing scribbles that leads to nonrepresentative drawings, but scribbles show the highest difference with other instructions for many metrics in dataset 1. Although it is possible to draw a figurative drawing using scribbles, no such cases were observed in our study. This is certainly because participants had toddler scribbles in mind when we gave the instruction. Observation of these differences shows that the process of drawing scribbles results in the rapid drawing of a small number of short sequences of relatively random lines.

Closer evaluation of the principal components analyses shows that some metrics obtained a high loading but were not found in the same dimensions in dataset 1 and dataset 2. This was the case for the minimum convex polygon, which does not show substantial differences between the instructions of dataset 1. We would expect the minimum convex polygon to be a proxy of representativeness or aestheticism by filling the screen, but toddlers are reported to fill the paper sheet when drawing (Kellogg, 1969; Matthews, 1984; Wolf, 1988). Indeed, the minimum convex polygon was also shown to be high with scribbles or other nonrepresentative drawings (see Fig. 3), which indicates that this metric cannot be used to understand the cognitive processes underlying drawing. The Gini and entropy indices, both measured on

temporal sequences of drawing, also belong to different dimensions in datasets 1 and 2. The Gini index is a measure of the inequality of temporal drawing sequences whilst entropy is a measure of the temporal uncertainty of drawing. Both measures could be developed to other aspects of the drawing (for instance, spatial Gini index and angle Gini index), but we think that our measures already covered these aspects and that adding other measures could overlap with the existing ones. Indeed, the duration of drawing sequences is linked to the lengths of the drawing lines for each object in the drawing. This can be seen in the different correlation charts, where temporal metrics are correlated to spatial ones. Given this uncertainty in the explanation of the dimensions of datasets 1 and 2, we preferred to remove these two metrics. However, PCA results did not change after this removal and the explained variance was higher. The Gini and entropy indices are increasingly used in different studies, despite a continuing debate about their interpretability (Ben-Naim, 2012; Leff, 2007; Lerman and Yitzhaki, 1984). Future works are required to assess their potential role in the domain of drawing and other behaviors. Moreover, the Gini index is ineffective when calculated on binary sequences and does not provide any new information. Perhaps considering the cumulative sum of drawing and non-drawing could lead to meaningful results.

After the different steps of selection of variables, we reached similar results between datasets 1 and 2 and explained almost 80% of variance. This result is important because it means that whatever the dataset and the given instructions, our method could be applied to analyze drawings. However, this method is only valuable if the dimensions of the PCA have a biological or psychological aspect.

Dimension 1 is composed of the drawing speed, the drawing distance and the μ MLE (see Table 1 for definitions). In movement ecology, the drawing speed is a proxy of

goal directedness—in other words, the intention of an animal to go to a specific place that it knows (Byrne et al., 2009; King and Sueur, 2011; Noser and Byrne, 2014; Sueur, 2011). The higher the motivation to go to a place where resources can be found, the higher the speed. In our drawing context, speed can be a proxy of intentionality and of mastering, meaning that participants who are familiar with what they are drawing do it faster. We observed this tendency in dataset 2, where the mean drawing speed of experts was 0.73 ± 0.37 compared with 0.56 ± 0.27 for naive participants. However, drawing speed is also high for scribbling toddlers or for someone who wants to fill the screen (or paper sheet) with one color, and in both of these cases, drawing speed is linked to drawing distance. This is what we obtained in our results (Tables 2 and 3; Fig. 3) with many participants coloring the starry sky blue or black. This adds detail without providing a better representativeness of the drawing. On the other hand, μMLE is negatively correlated to drawing speed and drawing distance, and is used in ecology to evaluate the efficiency of animal trajectories. In an environment with different food resources, an animal can move randomly or go directly to the resources area. In this case, movements are efficient and optimal and μMLE is high. We qualified these movements as a Lévy flight (or walk) (Reynolds, 2008; Viswanathan et al., 1999). In our drawing study, a high μMLE indicates an efficient drawing, i.e., it is representative, intentional and with few details (see Fig. 3). It is easy to recognize what the drawer wanted to draw, but drawing distance is low and indicates few details. If we could link an ability to this dimension 1, it would be efficiency. In this way, we suggest the dimension 1 to be close to the concept of efficiency. Indeed, efficiency can be defined as an ability to avoid wasting materials, energy, efforts, money, time, etc. Efficiency is different from effectiveness, which is the capability of producing a desired result or the ability to produce desired output

(Frøkjær et al., 2000; Marley, 2000). In our study, we suggest that effectiveness is representativeness, whilst efficiency is representativeness combined with few details (i.e., optimality). Efficiency might be illustrated in a sketch (Mihai & Hare, 2021; Xu et al., 2020) and by emoticons (Huang et al., 2008; Takahashi et al., 2017). Importantly, the scribbles made by adults showed higher values of dimension 1 than scribbles by toddlers. This is particularly due to higher drawing speed in adults (Table S2). The number of sequences is also higher for adults' scribbles, but the number of colors is lower. This may indicate that even if these drawings do not have an external representativeness, they may have an internal representativeness for adult participants. This also means that our analysis is able to show differences between drawings unable to see with our human eyes. Representative drawings do not mean figurative, in the way that an external eye sees something—objects or persons—drawn on the paper, but representative means that the drawer have something in mind to draw and this can be abstractive for instance.

The second PCA dimension is composed of the number of used colors and the mean colorimetric profile. Adding and diversifying colors facilitates the differentiation of objects in a drawing. When there are few details in a drawing, one color is enough to identify the object but when more details are present, the use of colors makes it easier to identify the different objects. This principle can be observed in Fig. 3a, for instance, with the drawings of the crab (one color) and the house (different colors to identify the flowers, the car, the butterfly, etc.). Colors facilitate the visual perception of objects and materials in our environment (Castelhano and Henderson, 2008; Witzel and Gegenfurtner, 2018). In our drawing datasets, this cognitive process is found to make drawings easier to interpret and increase their external

representativeness. We suggest the dimension 2 to represent “diversity” of colors in terms of number and panel.

Finally, the third PCA dimension is composed of the Hurst index and the number of sequences, both of which are temporal metrics. The number of sequences is directly dependent on the number of lines drawn (whatever their length), i.e., the number of forms that are either objects or components of an object. The Hurst index is a proxy of the temporal complexity of a drawing. It indicates how far the timeline of a sequence can predict another sequence. For instance, two drawings can contain the same number of sequences, but one will be considered as deterministic (small values of dimension 3, not complex, for analyses of dataset 1) if the duration of sequences is similar (because the drawn objects are all similar), whilst the other will be considered stochastic (high values of dimension 3) and more complex if the sequences cannot be predicted because they followed an unpredictable pattern (which is the intention when representing different objects). Such examples can be seen for instance in Fig. 3b and c (a drawing that resembles a rose window and the drawing with planets) for dataset 1. Here, higher values in dimension 3 seem to indicate higher anticipation and intention in drawing. We suggest dimension 3 to be close to the concept of sequentiality. When the complexity of the drawing increases to make something representative, the number of sequences and the stochasticity increase.

As expected, age influenced the three dimensions of drawing in dataset 2. The first dimension represents the μ MLE spatial fractal index, drawing distance, and drawing speed. It provides an insight into the representativeness and intentionality behind a drawing, even an abstract one; as well as the details needed to make a drawing figurative, such as the drawing distance. As such, the relationship between age and

dimension 1 is nonlinear and dimension 1 initially increases, then decreases with age. This trend has previously been found with the μ MLE spatial fractal metric (Martinet et al., 2021) but the dimension 1 from the PCA in this study gives more discriminative results. The dimension 1 increases in young children can be easily explained by the progressive development of more controlled and goal-oriented lines which often underlie the production of figurative drawings. Conversely, 3-year-old children are more motivated by motor pleasure alone when producing scribbles (Kellogg, 1969); therefore, dimension 1 is lowest for this age category. Thereafter, dimension 1 increases in 7- to 8-year-old children and decreases in 10-year-old children and adults. At 7 and 8 years of age, children draw all the parts of the object they have in mind without abstraction or unnecessary details (Baldy, 2011; Kellogg, 1969). Their primary goal is to be understood (external representativeness) with no aesthetic goals, which ensures that their drawings are more efficient. Adults' drawings appear more complex because of the compilation of numerous details. Furthermore, adult representations may be subject to other influences, such as social norms, which young children are not (Itskowitz et al., 1988); and our results in the two other dimensions confirm this. The dimension 2 represents the diversity and number of colors used. Young children are more motivated by the desire to play rather than drawing; thus, they try more colors and return the highest diversity values. Dimension 2 is also higher among novice adults due to their willingness to add more details. Children 7-9 years of age have the lowest diversity as their drawing efficiency is high. Dimension 3 represents the number of sequences and the temporal complexity of alternations; and it generally increased from the youngest children to the adults. Very young children draw with scribbles using few sequences, and are more deterministic than their older counterparts. Adults tend to add many details;

as such, they present the highest number of sequences and alternations between drawing and interruption resulting in a higher complexity. Therefore, the higher the dimension 3 in adults results in a lower dimension 1. Furthermore, the study of drawing's temporal components completes the spatial analysis and allows us to further understand the ontogenetic development of drawing behavior.

We also tested two other factors in dataset 2: the gender of the participants and the conditions of the drawing sessions (i.e., free drawing or self-portrait). Gender had no significant overall effects on the drawing dimensions, barring a minor influence on dimension 3. In a previous study, Martinet et al. (2021) found that gender had an effect on color use but not on the μ MLE spatial fractal metric. Other studies have shown that girls use more colors than boys in their drawings (Milne and Greenway, 1999; Wright and Black, 2013), but this mainly depends on the age of the children (Turgeon, 2008) and the instructions that they receive. Our results indicate that a higher number and more diverse spectrum of colors were used in free drawing versus self-portrait conditions, which may mitigate the gender effect. The open nature of free productions increased the use of colors, whereas the limitations imposed by self-portrait conditions lead individuals to use fewer elements when composing their drawing. (Martinet et al., 2021) also found that all age groups spent more time drawing under the free condition than under the self-portrait condition. This is an important bias to consider when thinking about research protocols, as the instructions given to participants appear to constrain their drawing process and may influence the results. Our study identifies three dimensions in drawing that we propose to represent the concepts of efficiency, diversity, and sequentiality. All three dimensions facilitate our understanding of drawing representativeness. This statement can be observed in Fig. 3, where we can easily determine the intentions

of participants, even if the drawing is abstract (i.e., there are no identifiable objects). The combination of these three dimensions allows us to judge the representativeness of a drawing even if it does not seem to represent anything for the observer and that the drawing is not figurative. In this way, we mean that if a person makes a drawing with (1) high speed and long lines, (2) many colors, and (3) many sequences, the drawing will get a high score and should be categorized as highly representative, in the way that these conditions cannot be met without a deep and strong reflection of the drawer. One condition taken apart can be the sign of motor development or locomotor pleasure, but our combined results allow us to determine if a drawer has an intention behind nonfigurative and abstract drawing. The fact to ask participants in this study abstract and nonfigurative drawings and that their dimension values, contrary to children's drawings, are closer to figurative drawings indicate the representativeness. The perspectives of this study are therefore noteworthy: we can use this method to evaluate the intentions behind a drawing that has no meaning for us, as adults with no psychopathologies. In this respect, we identify three perspectives: (1) We can study the ontogeny of drawing in children and identify with precision the premises of the different steps observed during the drawing learning process (action representation, romancing and guided elicitation). This part perspective is partly done in this study showing an effect on age on the three drawing dimensions but we need to go deeper in this ontogeny approach. (2) We can also extend the study to other species, particularly great apes, which are known to draw, and assess whether their drawing is motivated by internal representativeness (Martinet and Pelé, 2021). (3) Finally, the method can be extended to psychopathologies such as autism (Charman & Baron-Cohen, 1993; Jolley et al., 2013) and even certain emotional disorders (Desmet et al., 2021; Nolazco-

Flores et al., 2021), or simply be used to measure learning difficulties and creativity (Lee and Hobson, 2006; Urban, 2004). New technologies combined with new mathematical methods appear to be very useful and provide new possibilities to test mental states, intentions and emotions beyond these representations (Watanabe & Kuczaj, 2012). However, the meaning of each dimension has been assessed in visual terms only and is therefore not completely objective. New metrics could ultimately lead to the discover new cognitive dimensions and meanings, or reinforcing the discoveries of this study. Finally, the participants who have drawn for this study are all from France. To make these results universal, it could be useful to collect drawings from around the world.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s12110-022-09436-w>.

Acknowledgements

We thank the director and the teachers of the school for giving us access to their classrooms and showing interest in our research project. We are grateful to all the participants and to the parents of all the children, who accepted with enthusiasm to contribute to our study. Thanks also to Sarah Piquette, who provided help for the ethical components of this project.

Authors' Contribution

MP and CS supervised the study. MP and LM collected data. LM, BB and CS analysed data. CS wrote a first draft. All authors worked on the paper and approved the final version.

Funding

This project has received financial support from the CNRS through the MITI interdisciplinary programs and an IDEX Exploratory Research program from Strasbourg University.

Data Availability

The datasets generated during and/or analysed during the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.5387520>.

Declarations

Conflicts of Interest

The authors declare having no conflicts of interest.

Ethics Approval

The study was approved by the research ethical committee of Strasbourg University (Unistra/CER/2019-11).

Consent to Participate

Informed consent was obtained from all adult participants and from a parent or legal guardian for children. Informed consent for publication of identifying images in an online openaccess publication has been obtained too.

References

- Adi-Japha, E., Levin, I., & Solomon, S. (1998). Emergence of representation in drawing: The relation between kinematic and referential aspects. *Cognitive Development*, 13(1), 25-51
- Allaire, J. (2012). *RStudio: Integrated development environment for R*. Boston, MA, 770, 394.
- Bartumeus, F., Catalan, J., Fulco, U. L., Lyra, M. L., & Viswanathan, G. M. (2002). Optimizing the Encounter Rate in Biological Interactions: Lévy versus Brownian Strategies. *Physical Review Letters*, 88, 97901.
- Bartumeus, F., Catalan, J., Viswanathan, G. M., Raposo, E. P., & da Luz, M. G. E. (2008). The influence of turning angles on the success of non-oriented animal searches. *Journal of Theoretical Biology*, 252(1), 43-55. <https://doi.org/10.1016/j.jtbi.2008.01.009>
- Beltzung, B., Martinet, L., MacIntosh, A. J. J., Meyer, X., Hosselet, J., Pelé, M., & Sueur, C. (2021). To draw or not to draw: Understanding the temporal organization of drawing behaviour using fractal analyses. *bioRxiv*. <https://doi.org/10.1101/2021.08.29.458053>
- Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, 229(2), 209-220. <https://doi.org/10.1016/j.jtbi.2004.03.016>
- Ben-Naim, A. (2012). *Entropy and the second law: Interpretation and misinterpretations*. World Scientific Publishing Company.
- Bousquet, C. A. H., Petit, O., Arrivé, M., Robin, J.-P., & Sueur, C. (2015). Personality tests predict responses to a spatial-learning task in mallards, *Anas platyrhynchos*. *Animal Behaviour*, 110, 145-154.
- Brown, A. E., & De Bivort, B. (2018). Ethology as a physical science. *Nature Physics*, 14(7), 653-657.
- Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, 116(5), 472-480.
- Byrne, R. W., Noser, R., Bates, L. A., & Jupp, P. E. (2009). How did they get here from there? Detecting changes of direction in terrestrial ranging. *Animal Behaviour*, 77(3), 619-631. <https://doi.org/10.1016/j.anbehav.2008.11.014>
- Carl, P., Peterson, B. G., & Peterson, M. B. G. (2010). Package 'PerformanceAnalytics'. Retrieved March, 29, 2011.
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 660-675. <https://doi.org/10.1037/0096-1523.34.3.660>
- Charman, T., & Baron-Cohen, S. (1993). Drawing development in autism: The intellectual to visual realism shift. *British Journal of Developmental Psychology*, 11(2), 171-185. <https://doi.org/10.1111/j.2044-835X.1993.tb00596.x>
- Costall, A. (1995). The myth of the sensory core: The traditional versus the ecological approach to children's drawings. Cox, M. V. (2013). *Children's drawings*

of the human figure. Psychology Press. Cox, S. (2005). Intention and Meaning in Young Children's Drawing. *International Journal of Art & Design Education*, 24(2), 115-125. <https://doi.org/10.1111/j.1476-8070.2005.00432.x>

Debache, I., Bergouignan, A., Chaix, B., Sneekes, E., Thomas, F., & Sueur, C. (2019). Associations of Sensor-Derived Physical Behavior with Metabolic Health: A Compositional Analysis in the Record Multisensor Study. *International Journal of Environmental Research and Public Health*, 16(5), 741. <https://doi.org/10.3390/ijerph16050741>

DeLoache, J. S. (1991). *Symbolic Functioning in Very Young Children : Understanding of Pictures and Models*. 18.

Desmet, O. A., van Weerdenburg, M., Poelman, M., Hoogeveen, L., & Yang, Y. (2021). Validity and utility of the Test of Creative Thinking Drawing Production for Dutch adolescents. *Journal of Advanced Academics*, 1932202X21990099.

Dissanayake, E. (2001). *Homo aestheticus: Where art comes from and why*. University of Washington Press.

Ebeling, W., Molgedey, L., Kurths, J., & Schwarz, U. (2002). Entropy, Complexity, Predictability, and Data Analysis of Time Series and Letter Sequences. In A. Bunde, J. Kropp, & H. J. Schellnhuber (Éds.), *The Science of Disasters : Climate Disruptions, Heart Attacks, and Market Crashes* (p. 2-25). Springer. https://doi.org/10.1007/978-3-642-56257-0_1

Edwards, A., Phillips, R., Watkins, N., Freeman, M., Murphy, E., Afanasyev, V., Buldyrev, S., da Luz, M., Raposo, E., Stanley, H., & Viswanathan, G. (2007). Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer. *NATURE*, 449(7165), 1044-U5. <https://doi.org/10.1038/nature06199>

Freeman, N. H. (1993). *Drawing : Public instruments of representation*.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 345-352. <https://doi.org/10.1145/332040.332455>

Gardner, H. (1981). *Artful scribbles: The significance of children's drawings*. Gardner, H., & Wolf, D. (1987). *The symbolic products of early childhood*.

Gaston, K. J., & Fuller, R. A. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46(1), 1-9. <https://doi.org/10.1111/j.1365-2664.2008.01596.x>

Giraudoux, P., Giraudoux, M. P., & MASS, S. (2018). Package 'pgirmess'. Golomb, C. (2003). *The child's creation of a pictorial world*. Psychology Press.

Gurarie, E., Bracis, C., Delgado, M., Meckley, T. D., Kojola, I., & Wagner, C. M. (2016). What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of Animal Ecology*, 85(1), 69-84. <https://doi.org/10.1111/1365-2656.12379>

Handheld, G. P. S. (2020). 6 Estimating Travel Distance and Linearity of Primate Routes. *Spatial Analysis in Field Primatology: Applying GIS at Varying Scales*, 106.

- Holland, S. M. (2008). Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA, 30602-32501.
- Huang, A. H., Yen, D. C., & Zhang, X. (2008). Exploring the potential effects of emoticons. *Information & Management*, 45(7), 466-473.
- Jolley, R. P., O'Kelly, R., Barlow, C. M., & Jarrold, C. (2013). Expressive drawing ability in children with autism. *British Journal of Developmental Psychology*, 31(1), 143-149. <https://doi.org/10.1111/bjdp.12008>
- Kellogg, R. (1969). *Analyzing children's art*. McGraw-Hill Humanities, Social Sciences & World Languages.
- Kershenbaum, A. (2014). Entropy rate as a measure of animal vocal complexity. *Bioacoustics*, 23(3), 195-208.
- King, A. J., & Sueur, C. (2011). Where Next? Group Coordination and Collective Decision Making by Primates. *International Journal of Primatology*, 32(6), 1245-1267. <https://doi.org/10.1007/s10764-011-9526-7>
- Lee, A., & Hobson, R. P. (2006). Drawing self and others : How do children with autism differ from those with learning difficulties? *British Journal of Developmental Psychology*, 24(3), 547-565. <https://doi.org/10.1348/026151005X49881>
- Leff, H. S. (2007). Entropy, its language, and interpretation. *Foundations of physics*, 37(12), 1744-1766.
- Lerman, R. I., & Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters*, 15(3-4), 363-368.
- Luquet, G.-H. (1927). *Le dessin enfantin*. (Bibliothèque de psychologie de l'enfant et de pédagogie.).
- MacIntosh, A. (2014). The Fractal Primate : Interdisciplinary Science and the Math behind the Monkey. *Primate Research*, advpub. <https://doi.org/10.2354/psj.30.011>
- MacIntosh, A. J. J., Alados, C. L., & Huffman, M. A. (2011). Fractal analysis of behaviour in a wild primate : Behavioural complexity in health and disease. *Journal of The Royal Society Interface*. <https://doi.org/10.1098/rsif.2011.0049>
- Marley, J. E. (2000). Efficacy, effectiveness, efficiency. Martinet, L., & Pelé, M. (2020). Drawing in nonhuman primates : What we know and what remains to be investigated. *Journal of Comparative Psychology*.
- Martinet, L., Sueur, C., Hirata, S., Hosselet, J., Matsuzawa, T., & Pelé, M. (2021). New indices to characterize drawing behavior in humans (Homo sapiens) and chimpanzees (Pan troglodytes). *Scientific Reports*, 11(1), 3860. <https://doi.org/10.1038/s41598-021-83043-0>
- Matthews, J. (1984). Children drawing : Are young children really scribbling? *Early Child Development and Care*, 18(1-2), 1-39. <https://doi.org/10.1080/0300443840180101>
- Matthews, J. (1999). *The art of childhood and adolescence : The construction of meaning*. Taylor & Francis.
- Matthews, J. (2003). *Drawing and painting : Children and visual representation*. Sage.

- Meyer, X., MacIntosh, A. J., Chiaradia, A., Kato, A., Mattern, T., Sueur, C., & Ropert-Coudert, Y. (2017). Shallow divers, deep waters and the rise of behavioural stochasticity. *Marine Biology*, 164(6), 1-16.
- Meyer, X., MacIntosh, A. J., Chiaradia, A., Kato, A., Ramírez, F., Sueur, C., & Ropert-Coudert, Y. (2020). Oceanic thermal structure mediates dive sequences in a foraging seabird. *Ecology and evolution*, 10(13), 6610-6622.
- Mihai, D., & Hare, J. (2021). Differentiable Drawing and Sketching. arXiv preprint arXiv:2103.16194.
- Mitani, J., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour*, 45(4), 735-746. <https://doi.org/10.1006/anbe.1993.1088>
- Nilsen, E. B., Pedersen, S., & Linnell, J. D. C. (2008). Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions? *Ecological Research*, 23(3), 635-639. <https://doi.org/10.1007/s11284-007-0421-9>
- Nolazco-Flores, J. A., Faundez-Zanuy, M., Velázquez-Flores, O. A., Cordasco, G., & Esposito, A. (2021). Emotional State Recognition Performance Improvement on a Handwriting and Drawing Task. *IEEE Access*, 9, 28496-28504. <https://doi.org/10.1109/ACCESS.2021.3058443>
- Noser, R., & Byrne, R. W. (2014). Change point analysis of travel routes reveals novel insights into foraging strategies and cognitive maps of wild baboons. *American Journal of Primatology*, 76(5), 399-409. <https://doi.org/10.1002/ajp.22181>
- Papi, F., Liew, H. C., Luschi, P., & Chan, E. H. (1995). Long-range migratory travel of a green turtle tracked by satellite : Evidence for navigational ability in the open sea. *Marine Biology*, 122(2), 171-175. <https://doi.org/10.1007/BF00348929>
- Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., Cornilly, D., Hung, E., Lestel, M., & Balkissoon, K. (2018). Package ‘performanceanalytics’. R Team Cooperation.
- Piaget, J., & Inhelder, B. (2008). *The psychology of the child*. Basic books.
- Planckaert, J., Nicolis, S. C., Deneubourg, J.-L., Sueur, C., & Bles, O. (2019). A spatiotemporal analysis of the food dissemination process and the trophallactic network in the ant *Lasius niger*. *Scientific Reports*, 9(1), 1-11. <https://doi.org/10.1038/s41598-019-52019-6>
- Potts, J. R., Börger, L., Scantlebury, D. M., Bennett, N. C., Alagaili, A., & Wilson, R. P. (2018). Finding turning-points in ultra-high-resolution animal movement data. *Methods in Ecology and Evolution*, 9(10), 2091-2101. <https://doi.org/10.1111/2041-210X.13056>
- Racine, J. S. (2012). *RStudio : A platform-independent IDE for R and Sweave*.
- Revelle, W. (2011). An overview of the psych package. Department of Psychology Northwestern University. Accessed on March, 3(2012), 1-25.
- Revelle, W., & Revelle, M. W. (2015). Package ‘psych’. The Comprehensive R Archive Network. Reynolds, A. M. (2008). Optimal random Lévy-loop searching : New insights into the searching behaviours of central-place foragers. *EPL (Europhysics Letters)*,

- 82(2), 20001. <https://doi.org/10.1209/0295-5075/82/20001> Sueur, C. (2011). A Non-Lévy Random Walk in Chacma Baboons : What Does It Mean? PLoS ONE, 6(1), e16131. <https://doi.org/10.1371/journal.pone.0016131>
- Sueur, C., Briard, L., & Petit, O. (2011). Individual Analyses of Lévy Walk in Semi-Free Ranging Tonkean Macaques (*Macaca tonkeana*). PLoS ONE, 6(10), e26788. <https://doi.org/10.1371/journal.pone.0026788>
- Takahashi, K., Oishi, T., & Shimada, M. (2017). Is 😊 smiling ? Cross-cultural study on recognition of emoticon's emotion. *Journal of cross-cultural psychology*, 48(10), 1578-1586.
- Thomas, G. V., & Silk, A. M. (1990). *An introduction to the psychology of children's drawings*. New York University Press.
- Urban, K. K. (2004). Assessing creativity : The test for creative thinking-drawing production (TCT-DP) the concept, application, evaluation, and international studies. *Psychology Science*, 46(3), 387-397.
- Viblanc, V. A., Pasquaretta, C., Sueur, C., Boonstra, R., & Dobson, F. S. (2016). Aggression in Columbian ground squirrels : Relationships with age, kinship, energy allocation, and fitness. *Behavioral Ecology*, 27(6), 1716-1725. <https://doi.org/10.1093/beheco/arw098>
- Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, 401(6756), 911-914. <https://doi.org/10.1038/44831>
- Viswanathan, G., Raposo, E., & da Luz, M. (2008). Levy flights and superdiffusion in the context of biological encounters and random searches. *PHYSICS OF LIFE REVIEWS*, 5(3), 133-150. <https://doi.org/10.1016/j.plrev.2008.03.002>
- Watanabe, S. (2012). Animal Aesthetics from the Perspective of Comparative Cognition. In S. Watanabe & S. Kuczaj (Éds.), *Emotions of Animals and Humans* (p. 129-162). Springer Japan. https://doi.org/10.1007/978-4-431-54123-3_7
- Watanabe, S., & Kuczaj, S. (2012). *Emotions of Animals and Humans : Comparative Perspectives*. Springer Science & Business Media.
- Willats, J. (2005). *Making sense of children's drawings*. Psychology Press.
- Witzel, C., & Gegenfurtner, K. R. (2018). Color Perception : Objects, Constancy, and Categories. *Annual Review of Vision Science*, 4(1), 475-499. <https://doi.org/10.1146/annurev-vision091517-034231>
- Wolf, D. (1988). *Drawing the boundary : The development of distinct systems for spatial representation in young children*. *Spatial Cognition: Brain Bases and Development*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 231-245.
- Wolf, D., & Perry, M. D. (1988). From endpoints to repertoires : Some new conclusions about drawing development. *Journal of aesthetic Education*, 22(1), 17-34.
- Wolf, M., & Weissing, F. J. (2012). Animal personalities : Consequences for ecology and evolution. *Trends in Ecology & Evolution*, 27(8), 452-461. <https://doi.org/10.1016/j.tree.2012.05.001>

Xu, P., Hospedales, T. M., Yin, Q., Song, Y.-Z., Xiang, T., & Wang, L. (2020). Deep learning for freehand sketch: A survey and a toolbox. arXiv preprint arXiv:2001.02600.

Zeller, A. (2007). 'What's in a picture?' A comparison of drawings by apes and children. *Semiotica*, 2007(166), 181-214.

Chapitre 3

Chez l'humain : du gribouillage à la maîtrise



I. Article 4 : Quantifier le développement de dessin

Using convolutional neural networks predictions to study the development of drawing behavior in children

En préparation

Synthèse de l'Article 4

Afin de mieux comprendre l'émergence et le développement du dessin, différentes approches peuvent être envisagées, en considérant par exemple des représentations de figures humaines, ou encore le degré de figuration ou de symbolisme qui y figure. Seulement, définir et quantifier la présence de tels indices peut s'avérer complexe. Dans cet article, des dessins d'enfants de 2 à 10 ans sont comparés à des dessins d'adultes. Après avoir regroupé les dessins d'enfants dans de nouvelles catégories d'âge (2-3 ans, 3-4 ans, etc.), des réseaux de neurones convolutifs ont été entraînés afin de classer ces productions contre celles d'adultes. Plus l'écart d'âge entre les enfants et les adultes diminue, plus il sera difficile pour le modèle de classer correctement les dessins. Afin de quantifier la similarité entre les différentes classes d'âge, ces mêmes modèles sont ensuite utilisés dans un but prédictif.

Using convolutional neural networks predictions to study the development of drawing behavior in children

Benjamin Beltzung, Marie Pelé, Lison Martinet, Cédric Sueur

Introduction

In humans of all ages, drawing appears to be a powerful means of communication. Analyzing such productions can provide valuable insights into the human's mind. In psychology for example, this medium can help to highlight mental illnesses such as schizophrenia (Bozikas et al., 2004), depression (Eytan & Elkis-Abuhoff, 2013), or even to diagnose headaches (Stafstrom et al., 2002). More generally, numerous cognitive aspects can be investigated through drawing behavior. For example, the representation of human figure in drawings of humans vary depending on the culture. Binson et al. (2023) demonstrated that the body size and the body shape of the human figure were different between young adults from Israel and Thailand when asked to draw themselves. By analyzing drawings produced by children from different cultures, Restoy et al. (2022) demonstrated that the degree of individualism of the countries had an impact on the size and number of human figures in the drawings. But the culture is not the only factor that can influence the drawing behavior. By focusing on the development of the drawing behavior, San (1979) demonstrated that the use/choice of colors varies according to the age in children. The main colors in children's drawings before the age of 5 are red, blue and yellow, while colors such as black, white or orange start to appear at around 5 years old. Understanding how

the drawing behavior changes over growth is complex, and while many theories have been proposed, a lot of questions remain unanswered. Luquet (1927) suggested that the development of the drawing behavior can be seen as made up of 4 stages. The first stage, *fortuitous realism*, occurs when children lack intention to represent reality. At around the age of three, they find some resemblance between their drawing and marks from real world objects. Between the age of three and five, they enter the stage of *missed realism*, where they try to produce realistic drawings but do not have the abilities to do so. When children use their knowledge of the components of objects to represent them (around five to eight years old), they enter the *intellectual realism* stage, where drawings are figurative but some errors linked to transparency and perspective are still present. The last stage proposed by Luquet is *visual realism*, where these last errors are fixed.

While many other theories on the drawing behavior development exist, they may not be intrinsically exclusive, but may consider different concepts. While young children scribbles are usually described as meaningless, many reports have shown that motor pleasure may not be the sole cause of their drawing, but also an intention of representation (Freeman, 1993; Yamagata & Shimizu, 1997; Adi-Japha et al., 1998). Such evidences are classified in three categories: action representation, romancing and guided elicitation (Matthews, 1984; Adi-Japha, 1998). *Action representation* occurs when the drawing is associated with verbalization, for example when the child represents a moving object and produce the corresponding sound. When children are able to name their drawing but it remains challenging to interpret for another individual, they enter a phase of *romancing*. The *guided elicitation* phase occurs when children are unable to produce a figurative drawing on their own, but achieve

to fill an uncompleted drawing or produce a figurative drawing when helped by an adult.

As explained earlier, the human figure representation plays an important role in the drawing behavior. Baldy (2005) proposed a classification of the morphological development of the human figure based on 6 stages (the round and enumerated human figure, the tadpole, the intermediary, the conventional man (filiform and then tube-shaped) and eventually the outline.

While such theories are relevant, it may be difficult to visually interpret the corresponding stage for a given drawing. For example, if we try to classify drawings according to the categories proposed by Baldy, it may sometimes be difficult to know whether a drawing belongs to a category or another. Traditionally, drawings are analyzed by defining and extracting a set of features (main color, number of human figures...). While this approach can be insightful, two challenges arise. First, the amount of information contained in a drawing is substantial in nature, and using predefined features significantly limits the amount of extracted information. Then, as each of these feature focuses on a single aspect of the drawing, this approach does not consider the holistic aspect of drawings (i.e., considering drawing as a whole, and not only its different parts).. Moreover, studies in toddlers showed that their drawings may have some meanings even if they do not have figurative aspects on it (Cox, 1992 ; Freeman, 1993 ; Yamagata & Shimizu, 1997). For these reasons, such methods may not be sufficient to benefit from the information contained in the drawings to the fullest extent. A possible way to mitigate these issues is to ask children about what they intended to represent. However, they may not be directly conscious about the deep meaning of their drawing, and this could not be applied universally. For example, it could not be applied to scribbles drawn by very young

children who are not able to verbally communicate yet or children with pathologies that make them unable to communicate.

To minimize these biases, a potential candidate is deep learning. Over the past decade, important advances in deep learning models (neural networks) have been made when considering image, video or audio processing, substantially improving the predictive accuracy and outperforming state-of-the-art methods in many fields, such as system health management (Khan & Yairi, 2018), face recognition (Parkhi et al., 2015), or even speech recognition (Hassan et al., 2021). The most popular type of deep learning models is Convolutional Neural Networks (CNNs), which is known to provide a high accuracy for tasks involving image analyses. While CNNs architecture can vary according to the task, some key concepts remain. Different types of layers exist in CNN, such as convolutional layers. In convolutional layers, a convolution is applied on the image: a filter, representing a feature, slides over the image, and results in a feature map. Each value of the feature map is the degree of activation of the filter on the corresponding part of the image. Depending on the depth of the convolutional layers, filters can detect either low-level features (e.g., lines or curves) or high-level features (e.g., objects).

While such models usually provide a high accuracy, they are widely considered as black boxes with regard to the decision-making process (Lei et al., 2018 ; Schwartz-Ziv & Tishby, 2017), as it is not possible to straightly understand the process that led the model to predict a particular output. Indeed, the complexity and number of parameters, that can attain billion, make such models difficult to interpret. However, from this complexity also comes a strength. Deep learning models allow for analyzing every pixel of a given image, extracting a large amount of information contained in the images, and can therefore potentially grasp all the relevant

features. The features learning produced by neural networks is also complex, and allow for an objective feature representation.

Although deep learning has already been successfully used to analyze drawings (Beltzung et al., 2022, 2023; Chen et al., 2022; Philippsen & Nagai, 2020), even in children (Ali et al., 2022; Kissos et al., 2020), the development of the drawing behavior is, to our knowledge, only poorly analyzed (Long et al., 2018; Moon et al., 2022; Pysal et al., 2020; Yuan et al., 2020). In this paper, we first build and train a CNN by using Bayesian optimization to classify drawings according to the age of the drawers (i.e., 10 age categories, from children to adults). By using the same method, we then trained multiple models to classify children subcategories versus adults' drawings to compute the accuracy and predict drawings not belonging to the classes considered in the models.

Material and methods

a. Dataset

The data consist of 396 drawings produced by 150 children (from 2 to 10 years-old) and 43 adults (novice and experts) as shown in Table 1. All the participants were given a touchscreen displaying a white background and could choose among a panel of 10 colors by having access to an overlay on the bottom of the screen. Examples of drawings are shown in Figure 1. From 3 years-old to adults, the drawings were collected using the following protocol. Each participant was asked to produce two drawings, one under a free condition, where the participant was not given a particular instruction, and one under a self-portrait condition, where the drawer was asked to draw himself. Two categories of adults were defined prior to the data collection. First, a group of adults who had never taken drawing classes and did not

have drawing as a hobby, that will be here considered as novices. Then, an expert category, including art school students and professional illustrators. For more information about these data, please refer to (Martinet et al., 2021).



Figure 1. Examples of drawings. Each column respectively illustrates examples of 3 (a) and 7 (b) years-old, and experts' adults (c). The drawings shown in the first row are produced in free condition, while those shown in the second row are produced in self-portrait condition.

As 2 years-old participants were not verbally able to communicate and to understand the instructions, no particular task was given for these productions. For this reason, they will be considered as produced in free condition. In this category, 6 subjects participated and produced a total amount of 30 drawings, ranging from 3 to 8 drawings per individual (respectively 3, 4, 4, 5, 6, 8 drawings).

While drawing, participants could select an eraser in the bottom right-hand corner of the screen to erase part of their drawing. Non-intentional dots could appear if the drawer failed and pressed just next to the “erase” rectangle. For this reason, we inspected every drawing to remove these non-intentional dots. To do so, we defined a region including the 50% the most on the right, and the 33% the most on the bottom of the drawings, and removed dots when their distance with the closest dot outside

of the box was at least 350 pixels. The amount of drawings modified per category is presented in Table 1.

Age category	Drawing condition	Total number of drawings (number of modified drawings)	Number of modified drawings
2 years-old	Free	30	12
3 years-old	Self-portrait	20	17
	Free	19	
4 years-old	Self-portrait	20	15
	Free	20	
5 years-old	Self-portrait	20	11
	Free	20	
7 years-old	Self-portrait	21	10
	Free	22	
8 years-old	Self-portrait	18	5
	Free	17	
9 years-old	Self-portrait	21	12
	Free	20	
10 years-old	Self-portrait	22	13
	Free	22	
Novice	Self-portrait	21	23
	Free	21	
Expert	Self-portrait	21	10
	Free	21	

Table 1. Drawings distribution for each age category and drawing condition.

All drawings were of dimension 2732x2048, and were resized to a 224x224 square with 3 channels for the colors to conduct analyses.

a. Transfer learning from deep learning

To analyze the development of the drawing behavior through the age, we compared drawings from different age categories. To do so, we used deep learning, and more precisely transfer learning. Transfer learning is a method used in machine learning and consists in using the knowledge of an already trained model for another task (Torrey & Shavlik, 2010). This technique is particularly useful for small dataset,

which is the case here. For this reason, we also used data augmentation, more particularly horizontal flips, as this transformation does not distort the image and the result is still realistic. For this study, we used the architecture of VGG19 (Simonyan & Zisserman, 2015) pretrained on the ImageNet dataset (Deng et al., 2009). VGG models have already been widely used for drawings analyses (Long et al., 2018; Theodorus et al., 2020). VGG19 is a CNN consisting of 16 convolutional layers and 3 fully-connected layers. The last fully-connected layer was removed, as the ImageNet classes are not of interest in this study. By using VGG19 architecture and ImageNet weights, we first trained a model with 10 classes (i.e., the age categories). The architecture of all models is described below. Then, to refine these analyses, we considered, for children only, new age categories by grouping each class with the following one. For example, drawings produced by 3 and 4 years-old are gathered in a new category called “3-4 years-old”, drawings produced by 4 and 5 years-old will be in a new category called “4-5 years-old”, thus approximately doubling the number of drawings per category. This process was done up to and including 10 years-old. Novices and experts’ drawings were independently grouped in a new category simply called “adults”, in order to compare drawings from children to those of adults. To do so, we trained 7 models, each classifying drawings from a following pair of children’s categories against the adults’ drawings.

b. Bayesian optimization

For the training, we independently built models and tuned hyperparameters. Multiple strategies exist to find the best architecture and hyperparameters. For example, grid search consists in defining a set of vectors for each hyperparameter and training a model for every possible combination. Another possibility is random

search, taking into account the fact that hyperparameters may not all have the same impact.

Here, we used Bayesian optimization, a technique using the information of past evaluations to iteratively find the best combination from a given parameters space. The model is first trained with a set of parameters. This method then allows for using the prior knowledge (i.e., the knowledge of the past trials) to select a new set of parameters that are expected to improve the accuracy. The number of iterations of this process is defined beforehand.

The architecture of our model is based on VGG19, which takes a drawing with a (224,224,3) shape as input. Before flattening the result of the last convolution, we considered 3 pooling options through the Bayesian optimization process: average pooling, max pooling, or no pooling. After flattening, we considered a dropout layer through Bayesian optimization, with a value between 0 and 0.4 and a step (i.e., the smallest meaningful distance between two values) of 0.1, potentially followed by a Batch Normalization layer. Then, from 1 to 3 fully-connected layers are added, each containing from 32 to 512 units with a step of 32, each followed by a relu activation. The last fully-connected layer is eventually followed by the classification layer using sigmoid function and binary crossentropy loss. The learning rate is also optimized between $1e-5$ and $1e-1$ through log sampling with stochastic gradient descent (SGD) optimizer. The tuner is running on 50 trials (i.e., testing 50 different combinations of parameters) to minimize the validation loss. For each trial, the model is trained on 15 epochs, with a batch size varying from 16 to 64 with a step of 16. 80% of the data are used for the training and 20% for validation, with an early stopping if the validation accuracy did not improve for the last 3 epochs. Once all the trials have been evaluated, the optimal parameters are saved. The best architecture with the

best hyperparameters is then trained 10 times to compute the validation accuracy and finally the mean validation accuracy for each model. To assess the relative importance of colors, the same procedure was conducted after converting the images into grayscale.

c. Predictions matrices

For each model, we then computed the mean prediction for the age categories not belonging to the model's classes. As an output for each model, the sigmoid function returns a value between 0 and 1. For each model, children's drawings are labelled as 0 and adults' drawings as 1. Thus, if we consider 0.5 as the threshold, a drawing with an output value greater than 0.5 will be predicted as belonging to the adults' category (i.e., predicted as produced by an adult), while a value smaller than 0.5 will imply that the model classifies the drawing as belonging to the corresponding children category. This value can be interpreted as a measure of confidence of the model for the prediction. For example, if we consider the model classifying 3-4 years-old against adults, a value close to 0 means that the model is highly confident in predicting the drawing as belonging to the children category (here the 3-4 years old), and if the model is unsure about how to classify the drawing, the value will be close to 0.5. It is informative to look at the predictions for drawings belonging to the model's classes to compute the accuracy, but computing the predictions for drawings not belonging to any of the model's classes can also be of interest. For example, if we consider the model classifying 3-4 years-old against adults, computing the predictions for a drawing produced by a 10 years-old child will give insight into how the drawing will be considered by the model, quantitatively.

d. Grad-CAM

Deep learning models are known to be very efficient for images classification, however most models remain black boxes and disentangling features which played a role in the classification remains a hard task, making the interpretation of such models complex (Beltzung et al., 2023). In the present case, it is for example of interest to understand which features were discriminative. However, as a significant proportion of drawings are non-figurative, it is hard to answer this question. Instead, what can be done is looking at the regions of the images that played an important role in the classification. To do so, a possible approach is to use Grad-CAM (Selvaraju et al., 2017). For a given image, this algorithm displays a heatmap showing the important regions for a specific class. This is particularly interesting when computed on the predicted class, to understand why the model predicted the input image as belonging to this class. For a given class, the gradient is computed by using the activations of the last convolutional layer that captures high-level features. Here, the Grad-CAMs have been visualized on the validation data for every model.

Results

a. 10-classes model

The optimal model classifying drawings in 10 categories achieved an accuracy of 26.25%, compared to the 10% that would be achieved by a random model. The confusion matrix is shown in Figure 2.

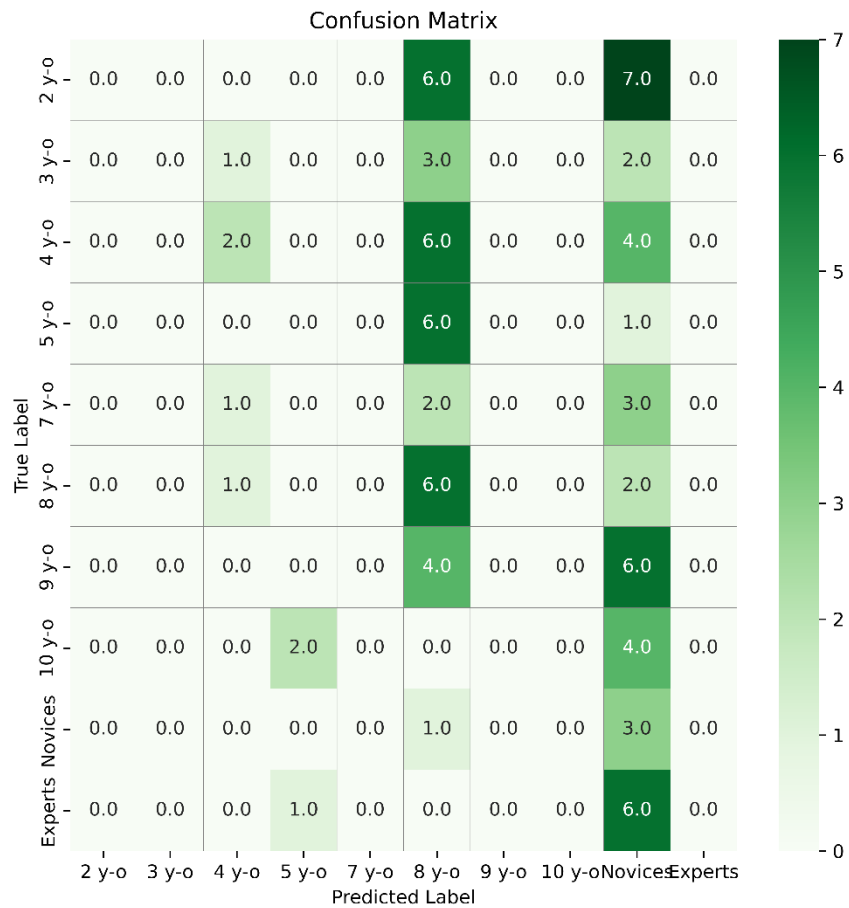


Figure 2. Confusion matrix for the 10-classes model.

a. Models by pair

To go further and have a better insight into the differences between age-categories, we trained models to classify adults’ drawings against children categories. The parameters of the optimal models that were selected through Bayesian optimization are shown in Table 3, as well as the mean accuracy (based on the validation accuracy calculated on the optimal model trained 10 times). The results for grayscale drawing models are shown in Table 4. The accuracy (computed 10 times) for RGB and grayscales models are shown in Figure 3.

Table 3. Optimal parameters and mean accuracy for every RGB model after the 50 Bayesian optimization trials.

Models Parameters	All categories	2-3 y-o	3-4 y-o	4-5 y-o	5-7 y-o	7-8 y-o	8-9 y-o	9-10 y-o
Pooling	Avg	None	Avg	Avg	None	Avg	Avg	Avg
Dropout	0.4	0.4	0.4	0.0	0.4	0.4	0.0	0.4
Batch Normalization	True	True	True	True	True	False	False	False
Number of FC layer(s)	3	3	1	1	1	1	3	1
FC1 layer size	32	288	32	512	32	512	512	512
FC2 layer size	512	512					512	
FC3 layer size	32	320					32	
Learning rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Batch size	16	64	16	64	64	16	64	64
Accuracy (mean over 10 iterations)		0.88	0.77	0.69	0.65	0.56	0.65	0.61

Table 4. Optimal parameters and mean accuracy for every grayscale model after the 50 Bayesian optimization trials.

Models Parameters	All categor ies	2-3 years -old	3-4 years -old	4-5 years -old	5-7 years -old	7-8 years -old	8-9 years -old	9-10 years -old
Pooling	Avg	Avg	Avg	Avg	Avg	Avg	Avg	Avg
Dropout	0.4	0.0	0.0	0.0	0.4	0.4	0.0	0.4
Batch Normalizatio n	True	False	False	True	True	True	False	True
Number of FC layer(s)	3	1	3	1	1	3	3	1
FC1 layer size	32	512	32	448	32	96	32	32
FC2 layer size	512		512			512	512	
FC3 layer size	32		512			448	512	
Learning rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Batch size	16	64	64	16	16	64	64	32
Accuracy (mean over 10 iterations)	0.26	0.93	0.64	0.78	0.71	0.63	0.68	0.68

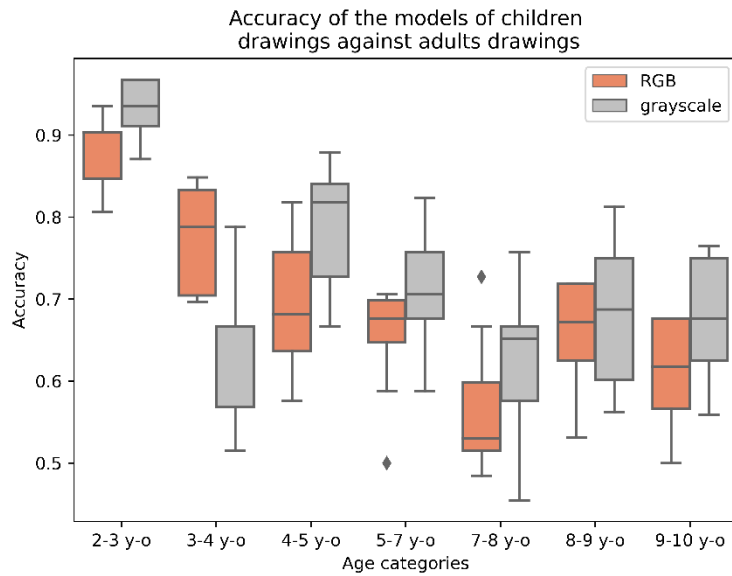


Figure 3. Accuracy for every model. Each model was trained 10 times. The accuracy was computed at each training.

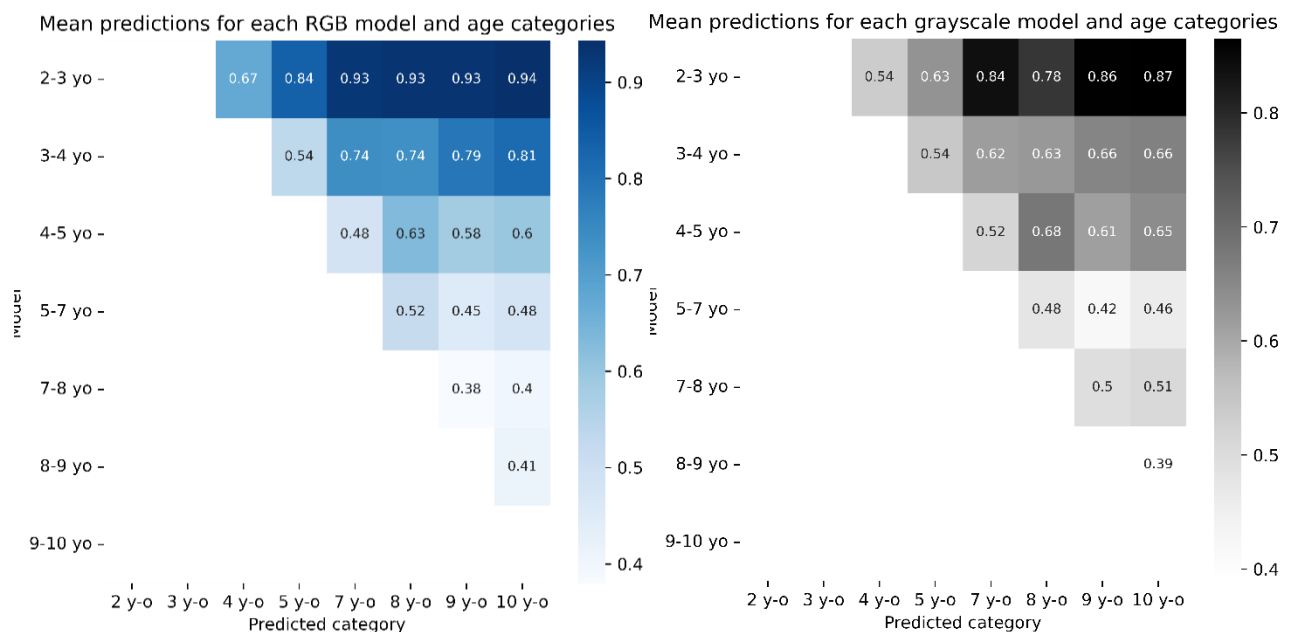


Figure 4. Mean prediction for each RGB (a) and grayscale (b) model and age categories.

For each model, the drawings belonging to the age categories between children and adults were predicted to compute the mean prediction for each model and class.

The results are presented in Figure 4 (a) for RGB and Figure 4 (b) for grayscale

models. The difference of prediction between RGB and grayscale models was eventually computed and is shown in Figure 5.

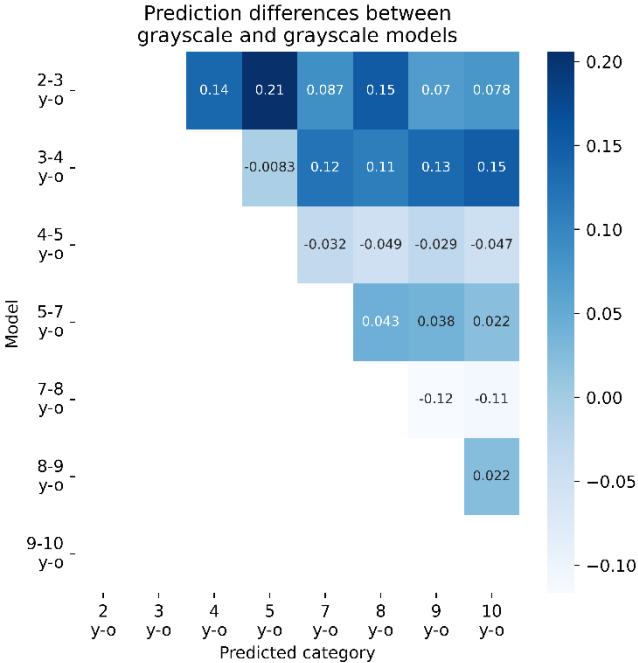


Figure 5. Difference of mean predictions between RGB and grayscale models.

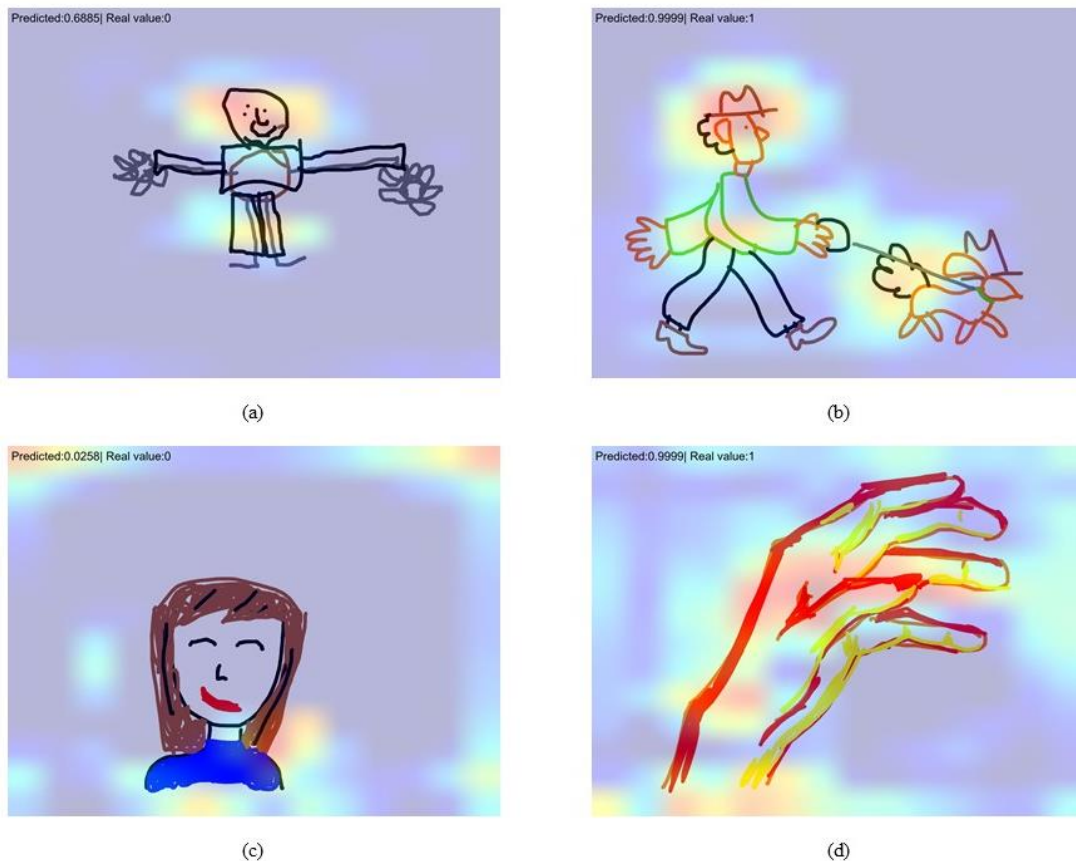


Figure 6. Examples of Grad-CAMs for drawings from the validation set of different models. Heatmap displayed in **a)** is from a 5 years-old (in the 4-5 years-old model). This drawing is wrongly classified as an adult drawing with a predicted value of 0.6885, although the heatmap seem to correctly detect and use the face as a feature. However, the drawing presented in **c)** is a 10 years-old drawing (in the 9-10 years-old model) and correctly classified, while the heatmap does not seem to make sense here. Heatmaps from correctly classified adults' drawings computed on the 2-3 years-old model sometimes seem to focus on specific features (**b)**) while sometimes the prediction remains very high although the heatmap does not seem to use any particular feature (**d)**).

Discussion

While it is well-known that drawings can be a window to the mind in children, it is not obvious whether the interpretation adults (experts or not) make of these drawings is consistent with the child's intention or not. Some indices may be prone to biases, specifically those that are not rigorously mathematically defined but

depend on the judgement of the observer. For this reason, it is important to develop new methods able to objectively analyze and decipher the information contained in drawings. Here, we propose a novel approach based on deep learning. By analyzing children's and adults' drawings and comparing different age categories, we demonstrate how deep learning can help at understanding the development of the drawing behavior. The first model, classifying drawings according to every age category achieves an accuracy of 26.25%, which is greater than the 10% that would be achieved by a random model. It is however difficult to assess which classes were the more similar (and thus mistaken), as most drawings were predicted to belonging to the 8 years-old and novices classes, according to the confusion matrix. After grouping children's categories by pairs, the most accurate model for RGB drawings was the one classifying 2 and 3 years-old against adults, with an accuracy higher than 85%. The accuracy decreased until the 7 and 8 years-old model, where it stabilized at approximatively 60%, which is quite interpretable and understandable as the older the children, the higher their drawings look like adults ones. However, removing the color information by using grayscale images surprisingly increased the accuracy of almost every model, except for the 3 and 4 years-old model, which accuracy lowered to about 60% compared to 80% for RGB models. This could be explained as color is not a factor increasing the understandability and representativeness of a drawing in all categories. However, for 3-4 yo children, color is a play tool contrary to other categories (Martinet et al. 2021). However, other factors intrinsic to the CNN model may intervene. First, the amount of drawings for every model is not high and may induce some variability when computing the accuracy, as the validation set is not very large. Moreover, it is possible that the number of trials used for this study is not sufficient and the accuracy for some classes may increase with the number of trials.

By using models to predict drawings from a category not belonging to the model's classes, we can assess if the age category is more similar to children's or adults' categories and quantify the degree of similarity.

In the prediction matrices, indeed, for the first row, every mean predictions are greater than 0.5, meaning that 4 to 10 years-old are more similar to adults than to 2-3 years-old. Mean predictions are increasing with the age of the predicted category, quantifying the development of the drawing skills. More interestingly, 4 years-old have a mean value of 0.67, showing that even 4 years-old drawings are more similar to adults' drawings than to 2-3 years-old drawings. This result demonstrates how important the development of drawing behavior is at young age. However, for the 3-4 years-old model, 5 years-old have a mean prediction of 0.54, which is very close to a random prediction. As this model does not significantly classify 5 years-old drawings in a category, one possible interpretation could be that 5 years-old drawing skill is an intermediate level between 3-4 years-old and adults. For the same model, the mean predictions increase from 7 to 10 years-old, but remain lower than the values found for the 2-3 years-old model. The 4-5 years-old model does not classify 7 years-old drawings as belonging more to children or adults' categories, and 8 to 10 years-old drawings are only slightly more classified as adults' drawings. The mean prediction values considerably decreased compared to the 3-4 years old model. A similar comment applies to the 5-7 years-old model, where the mean prediction drops to approximately 0.5. This demonstrates that 8 to 10 years-old drawings are as similar as 5-7 years-old drawings as adults' drawings. Compared to the 3 first models, the 2 last models do not classify the predicted categories as adults drawing, but as children drawings. Indeed, after reaching the 0.5 threshold, the mean prediction for 9 and 10 years-old for the 8-9 and 9-10 years-old models are

clearly below 0.5. These low values show that the drawing behavior development slows down between 7 and 10 years-old. For the first two rows, 4 and 5 years-old predictions are substantially lower than 7 to 10 years-old predictions. These results corroborate with previous findings (Long et al., 2018) showing that 7 to 10 years-old drawings were more similar to adults than to 4 to 6 years-old.

The heatmaps are particularly relevant for most drawings, but some ones are difficult to interpret as they were not classified correct. The heatmaps for some drawings (which had a very accurate prediction for some models) highlighted many regions that did not correspond to any understandable features. However, the model that led to the most visually interpretable heatmaps was the 2-3 years-old model, whose heatmaps were particularly understandable for adults (both novices and experts), in contrast to those of 2 and 3 years-old (i.e., non-figurative drawings). It is difficult to know whether heatmaps produced on non-figurative drawings are relevant, as we cannot be sure to read these drawings correctly.

The condition in which the drawings were produced, as well as the sex of the drawers, are not considered here. As it is known that these factors influence the spatial information contained in drawings (Martinet et al.), they may induce some variability in the analyses conducted in this paper. These variables are present in approximatively the same percentage for every age category, which may reduce their impact. Although the accuracies of the models are better than random, it is difficult to assess whether the omitted variables substantially impact the accuracy or not. This is particularly true for the model classifying drawings according to the age of the drawer, as this task may already be complex with the knowledge of these variables. As it has been discussed, the mean predictions matrix conveys a considerable amount of information. While it is informative to compare mean

predictions together and to the threshold 0.5, it is difficult to know how to quantitatively interpret an increase or decrease. For example, a value of 0.8 would not necessarily mean that the model is twice as confident as for a value of 0.4. Moreover, the accuracy of the model varies depending on the age categories. Models having a low accuracy will have more difficulties to discriminate children versus adults' drawings, which could also impact the mean prediction and its interpretation for other categories.

A new protocol where only one condition is given to the children but with more drawings per children would lead to have more interpretable results considering the age, as it would reduce the bias linked to the content. Concerning the models, increasing the number of trials and the hyperparameters space, as well as trying other architectures, could lead to better models and higher accuracy, which could particularly be of interest for the 10 classes model.

While looking at the accuracy and the confusion matrix for the 10 classes model is relevant, it would be interesting to compare them with human performing the same task. Asking humans to classify drawings according to the age categories could also give insight into the perception humans have of drawings and the development of this behavior.

Chapitre 4

L'intentionnalité chez les non- humains



I. Article 5 : Analyse exploratoire de dessins d'orang-outan

Using Artificial Intelligence to Analyze Non-Human Drawings: A First Step with Orangutan Productions

Publié dans *Animals*

Synthèse de l'Article 5

Comprendre l'origine évolutive du comportement de dessin implique d'analyser des dessins de primates non humains. Cependant, il est complexe d'estimer la place qu'occupe l'intentionnalité dans de telles productions, ainsi que les caractéristiques pertinentes à l'analyse de ces dernières. L'étude proposée dans cet article s'intéresse à la classification de plus de 1200 dessins de Molly, une femelle orang-outan, selon leur saison de production. Bien que cette tâche soit extrêmement complexe à réaliser à l'œil nu, l'utilisation de réseaux neuronaux convolutifs a permis d'atteindre une performance significative dans la classification de ces dessins. Dans un second temps, une analyse couche par couche a permis de mettre en lumière l'importance de différents niveaux de style et de contenu selon les saisons.

Using Artificial Intelligence to Analyze Non-Human Drawings: A First Step with Orangutan Productions

Benjamin Beltzung ^{1,*}, Marie Pelé ², Julien P. Renoult ³, Masaki Shimada ⁴ and Cédric Sueur ^{1,5}

- ¹ IPHC, Université de Strasbourg, CNRS, UMR 7178, F-67000 Strasbourg
- ² ANTHROPO-LAB, ETHICS EA 7446, Université Catholique de Lille, F-59000 Lille, France ;
marie.pele@univ-catholille.fr
- ³ CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France ;
julien.renoult@cefe.cnrs.fr
- ⁴ Department of Animal Sciences, Teikyo University of Science, 2525, Yatsusawa, Uenihara, Yamanashi, Japan; shimada@ntu.ac.jp
- ⁵ Institut Universitaire de France, 75231 Paris, France ;
cedric.sueur@iphc.cnrs.fr

Correspondence: benjamin.beltzung@iphc.cnrs.fr

Abstract: Drawings have been widely used as a window to the mind; as such, they can reveal some aspects of the cognitive and emotional worlds of other animals that can produce them. The study of non-human drawings, however, is limited by human perception, which can bias the methodology and interpretation of the results. Artificial intelligence can circumvent this issue by allowing automated, objective selection of features used to analyze drawings. In this study, we use artificial intelligence to investigate seasonal variations in drawings made by Molly, a female orangutan who produced more than 1299 drawings between 2006 and 2011 at the Tama Zoological Park in Japan. We train the VGG19 model to first classify the drawings according to the season in which they are produced. The results show that

deep learning is able to identify subtle but significant seasonal variations in Molly's drawings, with a classification accuracy of 41.6%. We use VGG19 to investigate the features that influence this seasonal variation. We analyze separate features, both simple and complex, related to color and patterning, and to drawing content and style. Content and style classification show maximum performance for moderately complex, highly complex, and holistic features, respectively. We also show that both color and patterning drive seasonal variation, with the latter being more important than the former. This study demonstrates how deep learning can be used to objectively analyze non-figurative drawings and calls for applications to non-primate species and scribbles made by human toddlers.

Keywords: primates; deep learning; drawing behavior; artificial intelligence; cognition

1. Introduction

Anthropocentric bias can arise when interpreting the meaning of what is depicted in drawings. Most studies on figurative drawings; nevertheless, non-figurative drawings can also have meanings, as found in young children's drawings (Quaglia et al., 2015). Moreover, drawings contain a large amount of information, and using a predefined set of handcrafted features limits the amount of information that can be extracted. This is particularly true when studying drawings made by non-human primates, as there may be an anthropocentric bias in choosing those features that are expected to be meaningful to us, even though they may be totally irrelevant to other species. Thus, predefined handcrafted features do not fully exploit the information content of drawings. In humans, a possible way to mitigate this issue is to ask the drawer what he or she intends to represent (Martinet et al., 2021). However, the drawer may not be directly aware of the deep meaning of the produced drawing. For example, the method cannot be applied to scribbles drawn by very young children who are not able to communicate verbally or to children who cannot communicate because of pathologies. Obviously, verbal communication about intentions is also not applicable to non-human animals that can draw, such as other primate species (Pelé et al., 2021).

In the present study, we used artificial intelligence to analyze the drawings of a female orangutan named Molly († in 2011), who produced 1299 drawings in her last five years as part of a behavioral enrichment program (Hanazuka et al., 2019) at the Tama Zoo in Japan. Previous investigations of Molly's drawings have already demonstrated that her drawings were influenced by her keeper's identity and daily external events, such as the birth of conspecifics (Hanazuka et al., 2019). The authors of (Pelé et al., 2021) investigated the differences among Molly's drawings

using traditional ethological methodologies (i.e., by selecting features and manually extracting them). These authors demonstrated the evolution of Molly's drawings over time and the influence of seasons. For instance, Molly used green color mostly during the summer and winter seasons and pink color in spring and autumn. How she drew lines also differed according to the season. However, because of the time required to extract features manually, these analyses were conducted on only 749 of the 1299 available drawings. Moreover, the authors focused on features typically used to study children's drawings (Kellogg, 1969), such as loops, circles, and fan patterns, which may have a different meaning for orangutans and humans or may even be totally irrelevant.

Deep learning can overcome these limitations. Deep learning is a set of artificial intelligence methods based on artificial neural networks (Jacob et al., 2021; Kuzovkin et al., 2018) and is now widely used in computer vision. Deep learning is an efficient way to replace or complement human expertise in images analyzed in complex tasks, such as classifying microscopy images (Buetti-Dinh et al., 2019) and diagnosing pathologies (Zhou et al., 2021). Deep learning does not require handcrafted features; it only requires labeled images. Convolutional neural networks (CNNs), the most common class of deep learning algorithms in image analysis, automatically learn and extract features that are most relevant for a given task (e.g., classification of images based on their labels). CNNs are composed of multiple layers, each containing filters, which are matrices that learn feature representations. The filters are initially randomly initialized and then optimized during the training. The complexity of the features increases with the depth of the layers, from low-level features (e.g., edges, circles) in shallow layers (i.e., close to the input) to high-level features (e.g., a nose, eyes) and entire objects (e.g., a face) in deeper layers (i.e.,

close to the output). Convoluting a filter with an image results in a new matrix, called a feature map. A feature map, also known as an activation map, represents the activation of the filter in different parts of the image. A high activation indicates the occurrence of a feature encoded by the filter in this part of the image. CNNs allow for high accuracy but have notoriously low interpretability. Nevertheless, multiple methods have recently been developed to better understand deep learning models, such as visualizing which parts of an image are important for its classification (Selvaraju et al., 2017).

Despite the power of artificial intelligence to analyze images, this method has rarely been applied to understand drawing behavior (Beltzung et al., s. d.). Thus far, most applications have exploited the high performance of deep learning in classification, for example, to recognize stroke representations (Wu et al., 2018) or to classify entire drawings according to the depicted object category (H. Zhang et al., 2016). However, the use of deep learning to investigate ontogeny or phylogeny has received little attention.

Here, we investigate multiple methods of using deep learning to decipher Molly's drawings. We first tested whether season could explain the variation among Molly's drawings by training a deep learning model to classify each drawing into one of the four seasons during which it was created. The results confirm the seasonal effect revealed in a previous study using handcrafted features (Pelé et al., 2021). To further understand this effect, we analyze how the information spreads within the network from shallow to deep layers. More precisely, we investigated the impact of feature complexity on classification accuracy. Finally, we leverage a particular type of artificial intelligence technique named 'style transfer' to dissociate features associated with drawing style from those describing the representational content,

and to analyze the relative importance of these two types of features in explaining seasonal variation in the predictions of the model.

2. Materials and Methods

2.1. Dataset

The data consist of 1299 drawings realized between 2006 and 2011 by Molly, who started drawing at approximately the age of 50 years and rarely had contact with conspecifics. She spent the morning in an enclosure (that could be inside or outside), and the afternoon in a restroom, where a crayon bucket was at her disposal so she could draw freely and whenever she wanted to. She spent approximately from 2 to 3 h every day, at the rate of 1-2 drawings per day. To end a session, Molly placed the drawing supplies on the floor. For more details on Molly and her drawings, please refer to (Hanazuka et al., 2019; Pelé et al., 2021). During this period, she was provided a piece of paperboard and colored crayons daily. As this activity was not specifically meant to investigate her drawing behavior, no rigorous protocol was applied. Thus, only few metadata were collected. To the best of our knowledge, the date of the production of the drawings is the only external source of information available. Each drawing was then categorized into one of the following four seasons based on the date: autumn, summer, spring, and winter (374, 284, 269, and 372 drawings, respectively). To fit the model discussed below, the drawings were reshaped to squares and resized to (224, 224, and 3). The dataset was split into a training set and a validation set with 907 and 392 images, respectively. For each set and season, the drawings were distributed as follows: 28.8% in autumn, 20.7% in spring, 21.9% in summer, and 28.6% in winter. Examples of drawings for each season are shown in Figure 1.



Figure 1. Examples of Molly's drawings. The season of each drawing, from the left to the right, is autumn, spring, summer, and winter. Credits goes to the Tama Zoological Park.

Apart from Molly's drawings, we considered two other datasets of animal drawings that should be easy to classify. We used these datasets to perform a sanity check on the performance of our artificial intelligence model. Pigcasso's dataset (Lefson, s. d.) includes 170 drawings of a male pig (*Sus scrofa*) living in South Africa who paints on a canvas by holding a brush in his snout. Pigcasso's drawing behavior was enhanced by positive reinforcement. The Pockets Warhol dataset (« Pockets Warhol dataset », s. d.) consists of 170 drawings made by a capuchin monkey (*Cebus capucinus*), who draws in a sanctuary in Canada by putting paint on his fingers and applying it to a canvas. Pockets Warhol started drawing as he was given children's paints to occupy himself.

2.2. Convolutional Neural Network (CNN)

As features pre-selected by humans may not grasp all the relevant information in drawings, we focused on deep learning, and more precisely, on CNN. The first CNN was trained on Molly's drawings to predict the season. We used the transfer learning method (Hussain et al., 2019) to circumvent the limitations of our relatively small dataset. Transfer learning consists of using an existing model that has already been trained on a particular task to perform another task. Following previous studies on drawings (Mukherjee & Rogers, 2020; Theodorus et al., 2020), we used the VGG19

architecture with ImageNet weights. VGG19 consists of 16 convolutional layers followed by 3 fully connected (FC) layers. We transferred the weights of the convolutional layers and randomly initialized the weights of the FC layers. The number of channels starts from 64 in the early layers, which captures simple elements such as lines, curves, and blobs (Qin et al., 2018), and grows to 512 in the deepest convolutional layers, allowing the network to learn a representation of complex objects. All convolution kernels were 3×3 . We tested different parameters and hyperparameters (e.g., learning rate and length of FC layers) to find the best model, the architecture of which is described as follows. The last convolutional layer from VGG19 is followed by an average pooling layer and a batch normalization layer. Then, a first FC layer of size 2048 with L2 regularization (strength of regularization of 0.1) is added, as well as a dropout layer with a rate of 0.2. The last layer is an FC layer of size 4, and each neuron corresponds to a season with softmax activation. We use a categorical cross-entropy loss function. To train the model, we used a stochastic gradient descent (SGD) optimizer with a learning rate of 0.1, and a batch size of 16. The best model was determined through early stopping by monitoring the validation set accuracy, and the training was stopped after three epochs with no improvement. As VGG19 is not specifically trained on drawings or paintings, we froze all convolutional blocks. To increase model accuracy, we also used data augmentation (horizontal and vertical flips). Illustration of the VGG19 architecture and examples of feature maps are shown in Figure 2.

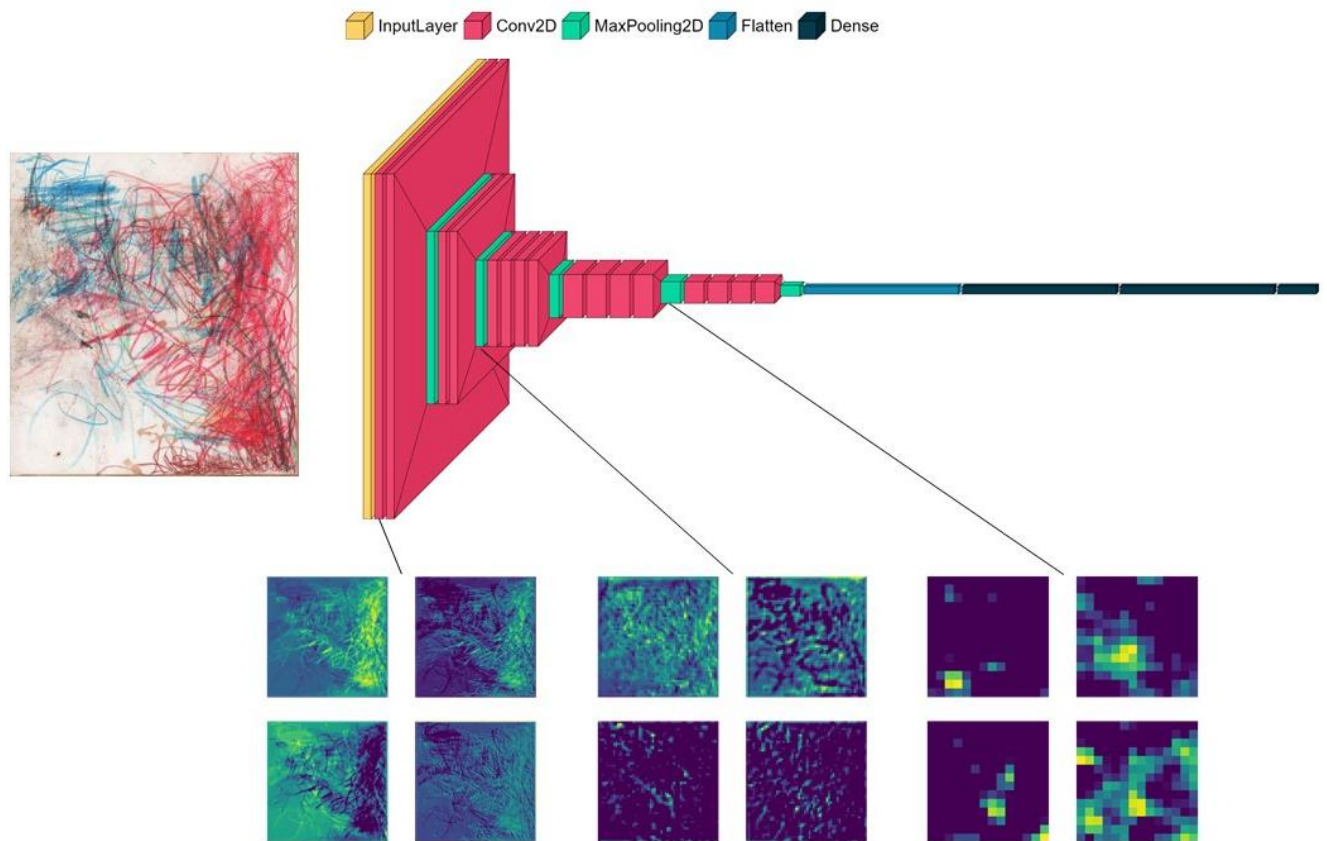


Figure 2. VGG19 model architecture and examples of feature maps (on the bottom) for a drawing. From left to right, each 2×2 square of feature maps are extracted from `block1_conv1`, `block3_conv1`, `block5_conv1`, respectively. The early blocks are in a high resolution, as the filters capture fine details, compared to deeper blocks, where the filters detect more general concepts.

2.3. Activations Classification

The previous classification of drawings according to season based on transfer learning was used as a baseline to compare other classification approaches intended to explore the role of different features in seasonal variation. Here, we focus on the features described by activations (Hulse et al., 2022). As shallow layers encode local information represented by simple features, and deeper layers encode global and more complex information, we first compared the ability of these different layers to classify seasons. For this purpose, each drawing was encoded using VGG19 (pretrained with ImageNet weights (Deng et al., 2009)), and for each layer separately, its activations were extracted, reduced, and used for classification. For

the first convolutional layer, for example, one image is represented by 64 matrices f_1, \dots, f_{64} of size 224×224 , corresponding to the 64 feature maps f_i of this layer. The maps are then flattened and concatenated into a single activation vector f . Then, principal component analysis is applied to the matrix containing the activation vector of every image, and only the first principal components (PCs) explaining a total of 80% of the variance are retained. The new vectors containing the retained PCs scores are then fed to a support vector machine (SVM) to classify images according to season. The same procedure is repeated for each convolutional layer, and for the first two fully connected layers, except that the activations were already vectorized, and thus did not require flattening. Accuracy was calculated through 10-fold cross-validation. Cross-validation consists of splitting the data into k -folds (here, $k = 10$) as follows: one-fold is used as the evaluation set and the $k-1$ other folds are used to train the SVM. This process is repeated for k folds; thus, each fold is used once for evaluation. For a given layer, the classification accuracy was calculated as the average of k scores.

For a given image, the length of f decreases with the depth of the layer. Thus, activation extraction does not require a significant amount of memory in deep layers. However, extracting the activations of every image requires a large amount of memory for early convolutional layers. In the first convolutional layer, f contains 3,211,264 activations. This became an issue when computing PCA for the first two convolutional layers. To address this issue, we reduce the size of f by randomly sampling 50% of its activation. To check whether this reduction preserved the information originally contained in the layers, we repeated the f reduction 50 times and compared the 50 means of the reduced f to the mean of the original f using a one-sample t -test. This test, performed for both the first and second convolutional

layers, was not significant, demonstrating that sampling 50% of the activations did not change their mean. Thus, for the first two convolutional layers, we considered half of the activations to compute PCA.

To understand the importance of colors in the seasonal variation of drawings, the previous SVM-based classification procedure was repeated using either RGB or grayscale images as inputs. The grayscale images were obtained using RGB channels and the National Television System Committee (NTSC) standard. To fit the VGG19 input shape, the new grayscale channel was repeated 3 times. For each layer, the RGB model was first compared to a dummy classifier to assess whether the RGB model was better than random classification. The RGB model was then compared with the grayscale model. To do so, we performed a 5×2 cv paired t-test (Dietterich, 1998). This test consists of comparing the accuracy of two models A and B by proceeding to five 2-fold cross-validations in a row, with the null hypothesis that both models have equal performance. The first accuracy is calculated using 50% of the data as training and 50% as testing, before reversing the roles. This was repeated five times for models A and B.

2.4. Gram Matrices Classification

Activation maps are often thought to encode information regarding the content of an image. To analyze whether Molly's drawing style varied across seasons, we repeated the classification procedures using Gram matrices of activations instead of the activations themselves (L. Gatys et al., 2015). For a given layer, the feature maps f_i are flattened and concatenated vertically to produce matrix F . The Gram matrix was then calculated as $G = F^T F$. Each value of G can be interpreted as the correlation between the two feature maps of the image. G was computed for every image before flattening. Each resulting vector represents a drawing, and these

vectors are fed to an SVM to classify the style of the images according to the season through a 10-fold cross-validation. The classification is repeated for every convolutional layer because the Gram matrices cannot be computed for the FC layers.

All the analyses were performed using Python.

3. Results

3.1. Classification through Transfer Learning

The model trained to classify drawings according to seasons achieved 41.6% accuracy on the test set. This accuracy is higher than that expected by random (approximately 29%, by always classifying drawings as the most common class). Drawings produced in the autumn (resp. spring) are often predicted to be produced in winter (resp. summer), and inversely, as shown by the confusion matrix (Figure 3). To assess the extent to which these two possible subgroups (autumn and winter vs. spring and summer) may differ, we retrained a model to classify these two subgroups, which achieved an accuracy of 66%.

For comparison, the same model that achieved 41.6% accuracy was trained to distinguish between Molly's drawings, Piggasso's drawings, and Pockets Warhol's drawings. The model achieved 93.6% accuracy in the 3-category classification task. As an additional comparison, the same model trained to classify Molly's drawings according to drawing coverage (binary classification: low coverage vs. high coverage) achieved 94.8% accuracy. These results show that our model is effectively able to classify drawings when the differences are large; thus, the significant but relatively low accuracy of the seasonal classification model is due to only subtle differences between seasons in Molly's drawings.

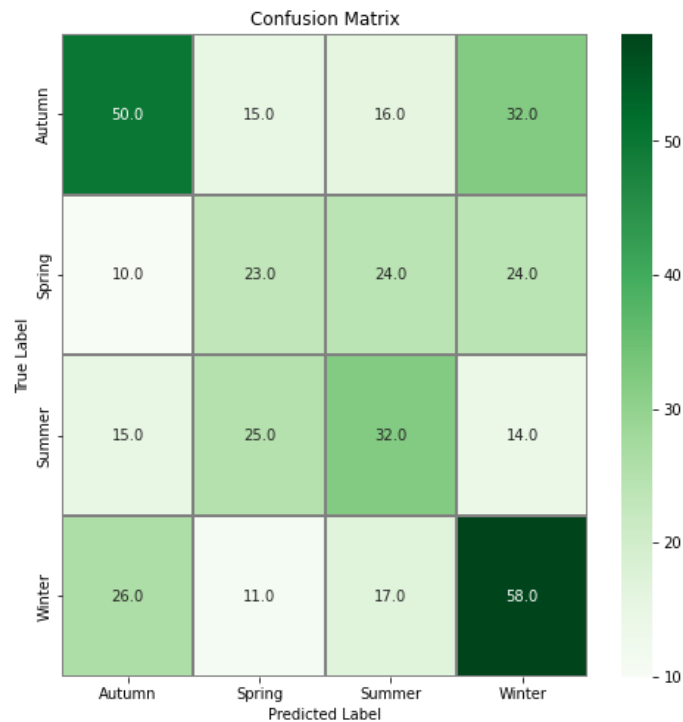


Figure 3. Confusion matrix of classification between seasons. For a given element, the value corresponds to the number of drawings predicted as the label matching with the column but belonging to the label matching with the row.

3.2. Layer-Wise Activation-Based and Gram Matrix-Based Classifications

The mean accuracy of the 10-fold cross-validation SVM at each layer for the RGB and grayscale images is shown in Figure 4. The accuracy of classification models based on activations of RGB images was compared to the accuracy of dummy models (always classifying images as belonging to the most frequent class) by conducting a 5×2 cv paired t-test. For every convolutional and fully connected layer, the model accuracy was different ($p < 0.05$) from the accuracy of the dummy models, showing that every layer encoded sufficient discriminative information to classify Molly’s drawings according to the season better than random. The slope of the linear regression model predicting the accuracy as a function of layer depth was not significantly different from zero ($p = 0.11$), showing that the accuracy did not improve through layers. Considering grayscale images, the accuracy of the models

differed ($p < 0.05$) from that of the dummy model, except for convolutional layers 1, 2, and 11. The slope for the grayscale images was positive and significantly different from zero ($p < 0.01$), indicating that the discriminative information contained in high-level features is not linked to colors only. The same test was performed to compare the accuracy between RGB and grayscale images (rather than the accuracy of dummy models) and revealed statistical differences only for layer FC2.

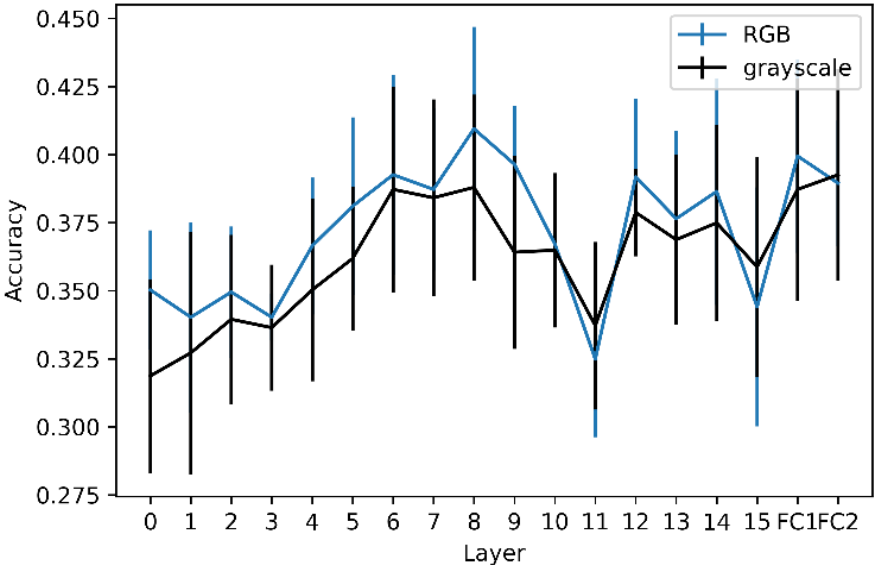


Figure 4. Mean accuracy of the 10-fold cross-validation for RGB images and grayscale images for each layer. The integer values on the x axis represent the convolutional layer on which the accuracy has been calculated, while FC1 and FC2, respectively, represent the first and second fully connected layers. The vertical bars represent the standard deviations.

The results of the Gram matrix-based classifications computed on RGB images are shown in Figure 5. Classification accuracy increased with layer depth ($p < 0.001$), indicating that complex-style features vary across seasons.

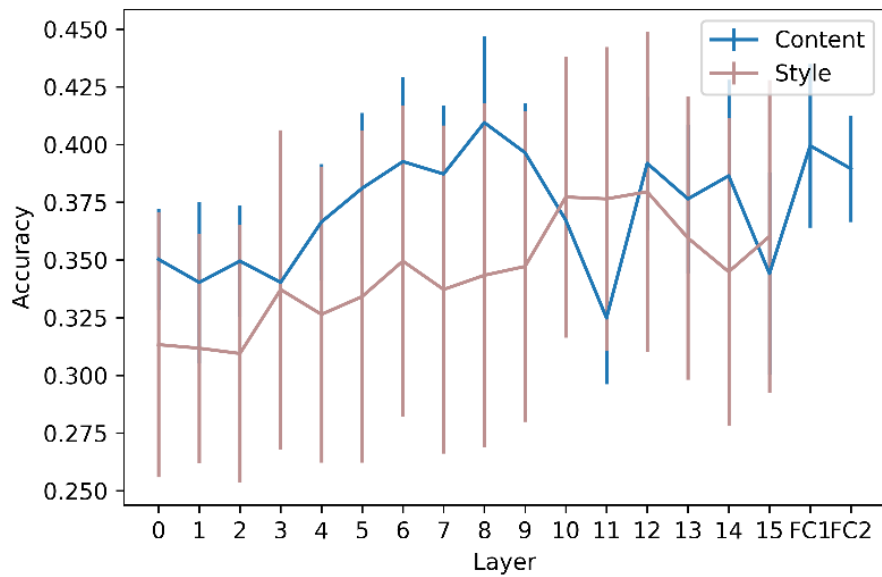


Figure 5. Accuracy of the models based on the Gram matrices (in brown) and on the RGB feature maps (in blue) for each layer. The vertical bars represent the standard deviations. The highest accuracy for Gram matrices classification is 37.9% for the 13th layer.

4. Discussion

The analysis of drawings is usually based on a limited number of features, which may fail to capture all the relevant information conveyed by this rich and generally meaningful communication and artistic medium. Moreover, the study of drawings is also limited by our subjective, human-based interpretation and understanding of what is represented, particularly for non-representative drawings, such as those produced by non-human primates. In this study, we proposed the use of deep learning models to overcome the limitations of traditional drawing analyses, as feature selection and extraction are not performed manually. Here, we focused on the classification of drawings of one female orangutan according to the season in which they were drawn. Previous analyses have shown differences based on the main color, where the main color is green in summer and winter, and pink in spring and autumn (Pelé et al., 2021). Thus, this study aimed to use deep learning to analyze these drawings and interpret the differences found in seasons and compare our results with

previous findings. To do so, we focused on convolutional neural networks (CNNs), the most common class of deep-learning models. By comparing the model accuracy with random predictions, we showed that the trained model was able to correctly classify drawings according to the season of production. Then, by exploring the information encoded by different levels of deep learning models, we could reveal how color, representational content, drawing style, and feature complexity contribute to seasonal variation in drawings.

Our general finding that deep learning methods can successfully encode visual information and detect variations among drawings is consistent with that of previous studies on abstract paintings. For example, (Bai & Guo, 2021) used a neural network to detect the correct orientation in such paintings, and (Lecoutre et al., 2017) used deep learning to recognize art styles in paintings, notably, abstract art. Our results extend these reports by showing that deep learning can discover structured variations that have often been considered scribbles with unpredictable variations in the literature (Papandreou, 2014).

Although the CNN model classification was better than random classification, the achieved accuracy remained relatively low (41.6%). This result could be explained by multiple reasons. The first interpretation comes from the confusion matrix, which shows that drawings produced in autumn are mostly misclassified as winter; conversely, drawings produced in winter are mostly misclassified as autumn. The same pattern was observed in spring and summer. These results are inconsistent with those of (Pelé et al., 2021), in which the suggested subgroups were summer and winter in autumn and spring. This may be because the CNN analyzes more complex features that were not considered in (Pelé et al., 2021). The relatively low accuracy of 41.6% may also be explained by the lack of a protocol when the drawings were

realized, but also by external events that occurred at Tama Zoological Park. For example, (Hanazuka et al., 2019) observed that the drawing produced by Molly on the day another orangutan gave birth was particularly red. They also observed that when elementary school students visited the zoo, the drawings were particularly rich in color and lines. It has also been demonstrated that the behavior of captive orangutans is directly linked to their familiarity with the keeper (Smith, 2014). The sex of the keeper and the people Molly has met could also influence her behavior, as was found for a female Japanese monkey (Koda et al., 1998), whose arterial blood pressure was different in the presence of men and women, and of strangers and caretakers. All these environmental factors likely added heterogeneity to Molly's drawings, which contributed to blurring the seasonal variation.

To further explore this seasonal variation, we considered that an image can be seen as an association of the following two components: content and style, and that these two components can be analyzed separately. Feature maps are commonly associated with the content of the image, where the content is 'the things represented or suggested in something written or created as art, or the ideas it communicates', according to the Cambridge Dictionary (« Cambridge Dictionary », s. d.). For a given image, the content will be encoded through all the layers and can be analyzed by extracting activations. The complexity of the content grasped in the activations increases with layer depth. The classification of the extracted activations at each layer was significantly better than random classification, showing the latent complexity of Molly's drawings, in which activations (i.e., content features) from low to high levels can be discriminated to a certain extent. The accuracy of activation extractions tends to increase from shallow layers to mid-layers, demonstrating the need for more complex and global content features than simple edges. However, the

slope of the accuracy as a function of layer depth was not significantly positive, demonstrating that high-level features linked to the content, describing the drawing as a whole, do not contain more relevant information for prediction than mid-level content features. However, this was not the case for grayscale images, where the slope was significantly different from zero. This difference between the analyses of color and grayscale versions of the drawings demonstrates that color information on seasonal variation in Molly's drawings is encoded more in the early layers than in the deeper layers. This is not surprising, considering that traditional analyses provide evidence for a difference in main colors depending on the season (Pelé et al., 2021). Indeed, the main color is a low-level feature, which should be captured in the early layers, which could partly explain the importance of colors in the early layers. It is important to note that, even if not significant, the accuracy of the RGB models seems higher than that of the grayscale models. This slight difference may be explained by the features linked to colors that contain information to predict the season of the drawings. Moreover, (Iigaya et al., 2021) demonstrated that shallow layers in VGG models encode information on the average hue and other low-level features, saturation, and other variables, whereas high-level features (i.e., deeper layers) can be linked to the concreteness and dynamics of the drawing. It is a complex task to disentangle high-level features that could have played a role in the predictions, but the average hue could be a proxy for the main color that was already analyzed in Molly's drawings (Pelé et al., 2021). The accuracies of the RGB and grayscale models were significantly different only for layer FC2. With our analyses, it is not possible to determine whether the relevant features for the RGB and grayscale models are the same. Moreover, differences between seasons are not only related to colors, as grayscale models are better than random models. As RGB models are not significantly

better than grayscale models (except for FC2), adding colors only slightly refines these differences, showing that the seasonal variation lies more in the shape of the strokes than their colors.

To complement the study of Molly's drawings, which describes what is depicted, we further investigated Molly's drawing style. To do so, we extracted Gram matrices that have been widely used in deep learning. However, how Gram matrices and styles are linked is not self-explanatory. Gram matrices in neural networks were first proposed to analyze the texture of images (L. Gatys et al., 2015). Here, the word 'style' also refers to the 'texture'. No consensus has been established on the meaning of these two words; however, according to (Sklansky, 1978), the region in an image has a constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying, or approximately periodic. The introduction of Gram matrices can be applied to analyze the style of an image dating back to (L. A. Gatys et al., 2016). In their study, the authors proposed decomposing artistic images into content and style, with the goal of transferring the style of an image onto another image whose content is kept unchanged. Our results suggest that the content represented by neuronal activation is not sufficient to extract all the latent information conveyed by Molly's drawings. First, the accuracy based on Gram matrices was higher than random for every layer, showing that Molly's drawing style varied across seasons. More importantly, style-based classification became more accurate when information from deeper layers was considered. This demonstrates that the layout of the drawings, that is, their global stylistic aspect, prevails over the local stylistic features. While the peak accuracy is reached in the intermediate layers in content analyses, it is reached in the deepest layers for style, which

highlights the importance of considering aspects, content, and style in drawing analysis.

The PCA applied to the Gram matrices showed that 80% of the variation could be retained with only two principal components for the first two convolutional layers. The corresponding scatter plots of PC1 versus PC2 revealed two possible clusters that did not appear to be season-dependent (Figure 6). With current data, it is difficult to identify the factors that explain these two clusters. They may be linked to the identity of the keeper or to various events that occurred at the Tama Zoological Park (Hanazuka et al., 2019). However, the corresponding scatter plots obtained from the activation vector (i.e., for content-based classification) also exhibited two clusters. These clusters become less distinct and less marked when looking at deeper layers and completely vanish in the deepest layers, suggesting that they are primarily rooted in low-level features such as edges (orientation and width).

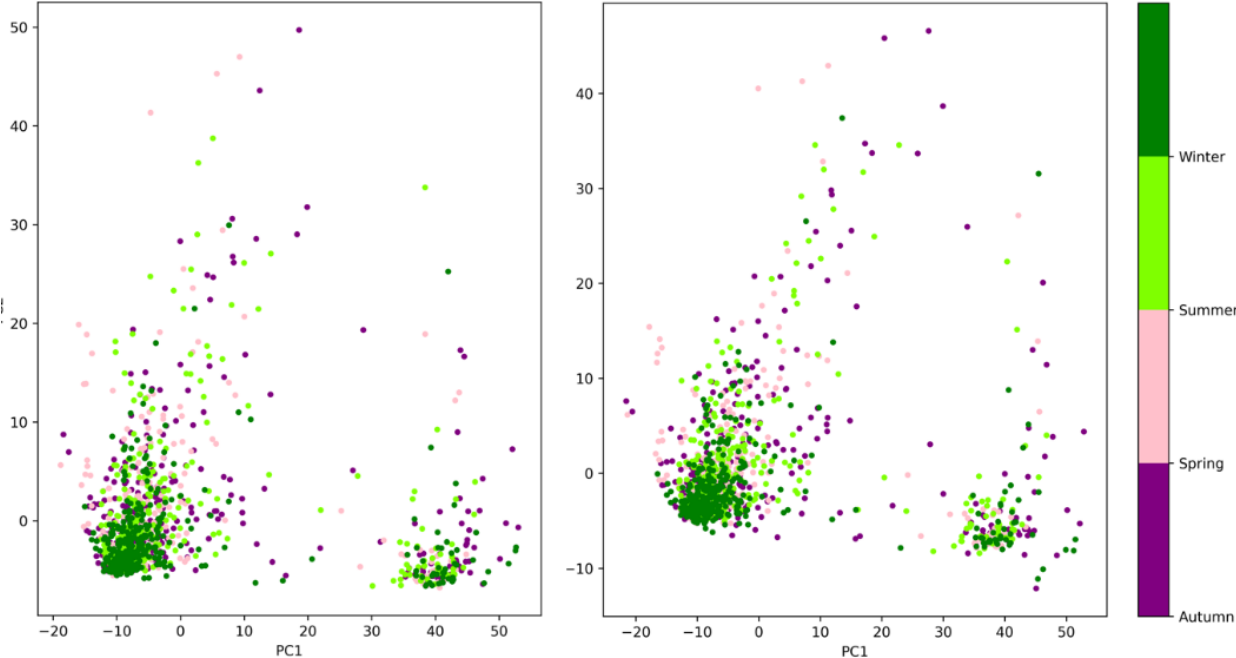


Figure 6. Scatter plot of the two PC of the Gram matrices of the first convolutional layer (on the left) and of the second (on the right), where each point is a drawing.

The approach of this study is based on the fact that biases can arise in analyses that require manual feature selection and extraction. However, deep learning reproduces biases such as race biases (Cavazos et al., 2021). Because we used transfer learning through the VGG19 architecture pretrained on a global classification task (the ImageNet dataset), the features learned by the model are generic. The choice of architecture is not crucial, as it has been shown that features from ImageNet pretrained VGG19 are shared when using other architectures (R. Zhang et al., 2018). Analyzing activations and Gram matrices does not require feature selection and extraction, thus being an objective methodology. Of course, the approach proposed here is still slightly anthropocentric when interpreting the results. Indeed, the concepts of content and style are human ones, and Molly may not perceive these concepts the same way as humans do. Deep learning and feature extraction partially allow for getting rid of anthropocentric biases as no feature selection or extraction involves humans. The authors of (Pelé et al., 2021) extracted 12 simple and objective features from orangutan drawings (e.g., coverage rate, distance to the center, number of loops, number of triangles), including Molly's ones, before performing a PCA, and the features the most correlated with the first axis were also highly linked to the content of the drawings whilst variables in the second correspond more to the style, demonstrating that even objective features will be linked to subjective ones. Whatever the approach, some anthropocentric bias linked to the production of the drawings will always remain. For example, the perceived affordance of a white canvas will be different for a human and a chimpanzee, and there will also be variability between individuals (Pelé, 2022). Regardless of these biases linked to the models or data collection procedures, the interpretation of the results will be performed by humans, making it impossible to completely erase these

anthropocentric biases. As Nagel wrote (Nagel, 1974), a human could never be able to know what it is to be another animal, but we think that artificial intelligence can help to decrease the number of biases.

5. Conclusions

Deep learning has great potential for improving our understanding of drawing behavior, both ontogenetically and phylogenetically. Such models can grasp the latent features present in abstract drawings, which is a time-consuming, laborious, and complex task for traditional analyses. The present case study on Molly's drawings demonstrates how the complexity of content and style, which is directly linked to Molly's drawing behavior, can be studied through a CNN, from shallow to deep layers. The layers of a CNN can be seen as the layers of a painting as follows: each layer contains partial information about the art at different levels of detail (such as colors, shapes, and contrast), and to understand a painting, one needs to combine these layers.

This study brings new advances in the understanding of drawings in non-human primates and can give highlights both on fundamental and more applied research. As regards non-verbal drawers, applying the current deep learning methodology to several ape species may lead to discoveries in the domain of evolutionary anthropology and comparative psychology. Moreover, analyzing drawings thanks to deep learning may help to assess neurodegenerative diseases in apes and enhance animal welfare. In humans, problems with languages (forgetting words, persons, etc.) allow targeting such problems but this cannot be performed with apes. Changes identified by AI in drawings with the age of individuals may help to identify such diseases.

Author Contributions

Conceptualization, B.B., M.P., J.P.R., and C.S.; data curation, S.M.; formal analysis, B.B. and J.P.R.; funding acquisition, M.P., S.M., and C.S.; investigation, B.B. and J.P.R.; methodology, B.B., M.P., J.P.R., and C.S.; project administration, M.P., S.M., and C.S.; software, B.B. and J.P.R.; supervision, C.S., M.P., and S.M.; validation, J.P.R., C.S.; visualization, B.B.; Writing—original draft, B.B.; Writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding

This project received financial support from the CNRS through the MITI interdisciplinary programs and from the University of Strasbourg through an IDEX Exploratory Research program. This study was partially supported by JSPS KAKENHI (grant number 20H01409).

Institutional Review Board Statement

The Tama Zoological Park Ethics Board approved this noninvasive behavioral study, which complied with the Code of Ethics of the Japanese Association of Zoos and Aquariums.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to Tama Zoological Park's request and copyright.

Acknowledgments

We are grateful to the Tama Zoological Park in Japan for providing the orangutan drawings.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the study design; collection, analyses, or interpretation of data; writing of the manuscript; or decision to publish the results.

References

- Quaglia, R.; Longobardi, C.; Iotti, N.O. Reconsidering the Scribbling Stage of Drawing: A New Perspective on Toddlers' Representational Processes. *Front. Psychol.* **2015**, *6*, 1227.
- Martinet, L.; Sueur, C.; Hirata, S.; Hosselet, J.; Matsuzawa, T.; Pelé, M. New Indices to Characterize Drawing Behavior in Humans (*Homo Sapiens*) and Chimpanzees (*Pan Troglodytes*). *Sci. Rep.* **2021**, *11*, 3860. <https://doi.org/10.1038/s41598-021-83043-0>.
- Pelé, M.; Thomas, G.; Liénard, A.; Eguchi, N.; Shimada, M.; Sueur, C. I Wanna Draw like You: Inter-and Intra-Individual Differences in Orang-Utan Drawings. *Animals* **2021**, *11*, 3202. <https://doi.org/10.3390/ani11113202>.
- Hanazuka, Y.; Kurotori, H.; Shimizu, M.; Midorikawa, A. The Effects of the Environment on the Drawings of an Extraordinarily Productive Orangutan (*Pongo pygmaeus*) Artist. *Front. Psychol.* **2019**, *10*, 2050.
- Kellogg, R. *Analyzing Children's Art*; National Press Books: Palo Alto, CA, USA, 1969.
- Jacob, G.; Pramod, R.T.; Katti, H.; Arun, S.P. Qualitative Similarities and Differences in Visual Object Representations between Brains and Deep Networks. *Nat. Commun.* **2021**, *12*, 1872. <https://doi.org/10.1038/s41467-021-22078-3>.
- Kuzovkin, I.; Vicente, R.; Petton, M.; Lachaux, J.-P.; Baciú, M.; Kahane, P.; Rheims, S.; Vidal, J.R.; Aru, J. Activations of Deep Convolutional Neural Networks Are Aligned with Gamma Band Activity of Human Visual Cortex. *Commun. Biol.* **2018**, *1*, 107.
- Buetti-Dinh, A.; Galli, V.; Bellenberg, S.; Ilie, O.; Herold, M.; Christel, S.; Boretska, M.; Pivkin, I.V.; Wilmes, P.; Sand, W.; et al. Deep Neural Networks Outperform Human Expert's Capacity in Characterizing Bioleaching Bacterial Biofilm Composition. *Biotechnol. Rep.* **2019**, *22*, e00321, <https://doi.org/10.1016/j.btre.2019.e00321>.
- Zhou, W.; Yang, Y.; Yu, C.; Liu, J.; Duan, X.; Weng, Z.; Chen, D.; Liang, Q.; Fang, Q.; Zhou, J.; et al. Ensembled Deep Learning Model Outperforms Human Experts in Diagnosing Biliary Atresia from Sonographic Gallbladder Images. *Nat. Commun.* **2021**, *12*, 1259, <https://doi.org/10.1038/s41467-021-21466-z>.
- Selvaraju, R.R.; Das, A.; Vedantam, R.; Cogswell, M.; Parikh, D.; Batra, D. Grad-CAM: Why Did You Say That? *arXiv* **2017**, arXiv:1611.07450.
- Beltzung, B., Pelé, M., Renoult, J. P., Shimada, M., & Sueur, C. (2022). Using Artificial Intelligence to Analyze Non-Human Drawings: A First Step with Orangutan Productions. *Animals*, *12*(20), 2761.
- Wu, X.; Qi, Y.; Liu, J.; Yang, J. Sketchsegnet: A Rnn Model for Labeling Sketch Strokes. In Proceedings of the 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), Aalborg, Denmark, 17-20 September 2018; pp. 1-6.
- Zhang, H.; Liu, S.; Zhang, C.; Ren, W.; Wang, R.; Cao, X. Sketchnet: Sketch Classification with Web Images. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016; pp. 1105-1113.
- Lefson, J. Pigcasso dataset, pigcasso.org
- Pockets Warhol dataset, pocketswarhol.blogspot.com
- Hussain, M.; Bird, J.J.; Faria, D.R. A Study on CNN Transfer Learning for Image Classification. In *Proceedings of the Advances in Computational Intelligence Systems*; Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 191-202.

Mukherjee, K.; Rogers, T.T. Finding Meaning in Simple Sketches: How Do Humans and Deep Networks Compare? *J. Vis.* **2020**, *20*, 1026-1026. <https://doi.org/10.1167/jov.20.11.1026>.

Theodorus, A.; Nauta, M.; Seifert, C. Evaluating CNN Interpretability on Sketch Classification. In Proceedings of the Twelfth International Conference on Machine Vision (ICMV 2019), Amsterdam, The Netherlands, 16-18 November 2019; Volume 11433, pp. 475-482.

Qin, Z.; Yu, F.; Liu, C.; Chen, X. How Convolutional Neural Network See the World—A Survey of Convolutional Neural Network Visualization Methods. *arXiv* **2018**, arXiv:1804.11191. <https://doi.org/10.3934/mfc.2018008>.

Hulse, S.V.; Renoult, J.P.; Mendelson, T.C. Using Deep Neural Networks to Model Similarity between Visual Patterns: Application to Fish Sexual Signals. *Ecol. Inform.* **2022**, *67*, 101486, <https://doi.org/10.1016/j.ecoinf.2021.101486>.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. Imagenet : A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20-25 June 2009; pp. 248-255.

Dietterich, T.G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.* **1998**, *10*, 1895-1923. <https://doi.org/10.1162/089976698300017197>.

Gatys, L.; Ecker, A.S.; Bethge, M. Texture Synthesis Using Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7-12 December 2015; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.

Bai, R.; Guo, X. Automatic Orientation Detection of Abstract Painting. *Knowl.-Based Syst.* **2021**, *227*, 107240. <https://doi.org/10.1016/j.knosys.2021.107240>.

Lecoutre, A.; Negrevergne, B.; Yger, F. Recognizing Art Style Automatically in Painting with Deep Learning. In Proceedings of the Ninth Asian Conference on Machine Learning, Seoul, Korea, 15-17 November 2017; pp. 327-342.

Papandreou, M. Communicating and Thinking Through Drawing Activity in Early Childhood. *J. Res. Child. Educ.* **2014**, *28*, 85-100. <https://doi.org/10.1080/02568543.2013.851131>.

Smith, J.J. Human-Animal Relationships in Zoo-Housed Orangutans (*P. Abellii*) and Gorillas (*G. g. Gorilla*): The Effects of Familiarity. *Am. J. Primatol.* **2014**, *76*, 942-955. <https://doi.org/10.1002/ajp.22280>.

Koda, N.; Machida, S.; Goto, S.; Nakamichi, M.; Itoigawa, N.; Minami, T. Cardiac and Behavioral Responses to Humans in an Adult Female Japanese Monkey (*Macaca Fuscata*). *Anthrozoös* **1998**, *11*, 74-78. <https://doi.org/10.2752/089279398787000797>.

Cambridge Dictionary. dictionary.cambridge.org

Igaya, K.; Yi, S.; Wahle, I.A.; Tanwisuth, K.; O'Doherty, J.P. 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>.

Sklansky, J. Image Segmentation and Feature Extraction. *IEEE Trans. Syst. Man Cybern.* **1978**, *8*, 237-247. <https://doi.org/10.1109/TSMC.1978.4309944>.

Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016; pp. 2414-2423.

Cavazos, J.G.; Phillips, P.J.; Castillo, C.D.; O'Toole, A.J. Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias? *IEEE*

Trans. Biom. Behav. Identity Sci. **2021**, *3*, 101-111.
<https://doi.org/10.1109/TBIOM.2020.3027269>.

Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-22 June 2018; pp. 586-595.

Pelé, M. Au Bout d'un Crayon, Un Singe. In *Mondes Animaux-Mondes Artistes*; Renoue, M., Pelé, M., Baratay, É, Eds. ; Presses Universitaires de Valenciennes : Valenciennes, France, 2022 ; *in press*.

Nagel, T. What Is It like to Be a Bat? *Philos. Rev.* **1974**, *83*, 435-450.

II. Article 6 : Les animaux et les machines sont-ils des artistes ?

Vers une création picturale non humaine : quel degré d'auteurisation pour les animaux et les machines ?

Publié dans *Polygraphe(s)*

Synthèse de l'Article 6

Ces dernières années, les modèles génératifs ont connu une croissance exponentielle, et il est aujourd'hui très simple d'utiliser ces derniers pour générer une peinture ou une photo créée de toutes pièces. Afin d'atteindre cet objectif, les modèles génératifs sont entraînés sur des jeux de données de peinture ou de photos. La question se pose alors de savoir qui considérer comme auteur des images générées. Mais cette question n'est pas inhérente à l'intelligence artificielle. En effet, il est difficile de savoir si nous devrions considérer les chimpanzés comme auteurs de leurs dessins. Afin de répondre à cette question, cet article propose des pistes de réflexion mêlant philosophie, éthologie, éthique et droit.

Vers une création picturale non humaine : quel degré d'auteurisation pour les animaux et les machines ?

Cédric Sueur^{1,2,3}, Benjamin Beltzung¹, Marie Pelé⁴

1 : Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

2 : Institut Universitaire de France, Paris, France

3 : CEERE, Centre Européen d'Enseignement et de Recherche en Éthique, Strasbourg, France

4 : Anthro-Lab, ETHICS EA7446, Université Catholique de Lille, Lille, France

cedric.sueur@iphc.cnrs.fr ; benjamin.beltzung@iphc.cnrs.fr ; marie.pele@univ-catholille.fr

Résumé : Imaginez un ver de terre se déplacer sur une feuille de papier en laissant une trace de peinture derrière lui. Imaginez maintenant un chimpanzé peindre sur une feuille de papier à l'aide d'un pinceau. Envisagez-vous un processus créatif identique chez ces deux animaux ? Ces deux animaux peuvent-ils être désignés comme auteurs de leur création ? Si un seul, lequel ? L'auteurisation d'un animal par une création dépend certainement de son intention de créer, c'est-à-dire de son agentivité couplée à la conscience qu'il possède de cette dernière. Ces notions sont très discutées dans les domaines de l'éthique animale et du droit animalier, tirant leurs éléments réflexifs de la biologie et de l'éthologie cognitive en particulier. Imaginez maintenant un robot doté d'une intelligence artificielle en train de dessiner ou de peindre. Comment considéreriez-vous de telles créations ? Nous discuterons ici des éléments issus de l'éthologie, de l'éthique et du droit permettant d'établir des bases de réflexion sur les concepts d'auteurisation et d'appropriation de créations graphiques par des entités non humaines.

Mots clés : intelligence artificielle, éthologie, éthique animale, droit animalier, conscience

Abstract: Imagine an earthworm moving around on a sheet of paper, leaving a trail of paint behind. Now imagine a chimpanzee painting on a sheet of paper using a brush. Do you envision an identical creative process in these two animals? Can these two animals be designated as the authors of their creation? If only one, which one? The authorization of a creation by an animal certainly depends on its intention to create, that is, on its agency coupled with its awareness of the latter. These notions are much discussed in the fields of animal ethics and animal law, drawing their reflective elements from biology and cognitive ethology in particular. Now imagine a robot with artificial intelligence drawing or painting. How would you consider such creations? We will discuss here elements from ethology, ethics and law allowing to establish a basis for reflection on the concepts of authorization and appropriation of graphic creations by non-human entities.

Keywords: artificial intelligence, animal behaviour, animal ethics, animal law, consciousness

Introduction

« Somme toute, l'artiste n'est pas seul à accomplir l'acte de création, car le spectateur établit le contact de l'œuvre avec le monde extérieur en déchiffrant et en interprétant ses qualifications profondes et par là ajoute sa propre contribution au processus créatif. Cette contribution est encore plus évidente lorsque la postérité prononce son verdict et réhabilite des artistes oubliés. »
Marcel Duchamp, *Le processus créatif* (1957)

Au même titre que l'utilisation d'outils, que le rire, la musique ou encore que les mathématiques, l'être humain a longtemps considéré l'art, notamment graphique, comme lui étant propre. Pourtant, loin d'être des Picasso ou des Monet aux yeux des êtres humains, certains animaux n'en présentent pas moins des comportements graphiques que nous pourrions qualifier d'artistiques (Watanabe, 2012). En effet, comme questionné par Marcel Duchamp, l'art ne naît-il pas à la fois de l'acte du créateur et du regard du spectateur ? En Océanie, les oiseaux à berceaux mâles forment des structures de branches complexes pour attirer les femelles et ainsi accéder à la reproduction (Diamond, 1986) : des cailloux, des graines ou des feuilles, le plus souvent colorés, sont déposés par les oiseaux à l'entrée du berceau afin d'en changer la symétrie, l'esthétique et la perspective. Une graduation de ces différents éléments, des plus grands aux plus petits ou des plus sombres aux plus clairs, peut ainsi être respectée par les oiseaux pour affiner leur décorum. Également dans le but de séduire les femelles, les mâles poissons-ballons sculptent des rosaces dans le sable avec leur bouche (Kawase *et al.*, 2013). Ces structures sableuses rappellent les dessins complexes réalisés sur le sable par les Ni-Vanuatu, ethnies mélanésiennes originaires du Vanuatu (Zagala, 2004). Un dernier exemple de compositions complexes chez les animaux que nous pourrions apprécier artistiquement est sans nul doute les toiles d'araignées (Krink et Vollrath, 1997). La réalisation de ces structures complexes sous-tend des processus d'apprentissage élaborés tout comme

des règles simples qui permettent d'augmenter l'efficacité souhaitée de la structure produite. Tels des algorithmes, la sélection naturelle, comme la sélection sexuelle, ont façonné les processus de réalisation de ces créations animales qui parfois nous touchent et nous émeuvent. Mais ces créations animales entrent-elles dans le domaine de l'art ? La réponse à cette question semble difficile à trouver. À un niveau moindre, nous pouvons cependant nous demander si ces animaux créateurs peuvent être considérés comme auteurs de leurs productions picturales. Tout comme l'art, la création picturale suppose une notion d'intentionnalité de la part de son auteur, un but à atteindre (Beardsley, 1970). Nous pouvons dès lors nous questionner sur les niveaux d'intentionnalité et de conscience d'action de ces animaux. Cette réflexion peut également se prolonger aux intelligences artificielles (Mikalonytė, Kneer, 2021) dont les créations, parfois indistinguables de créations humaines, se monnayent désormais plusieurs centaines de milliers d'euros (Doherty, 2019). Parallèlement, des animaux aux capacités cognitives proches de celles de l'être humain dessinent et peignent quand on leur en donne la possibilité (pour une revue chez les grands singes, voir Martinet et Pelé 2020). Leurs créations, souvent comparées à des gribouillages d'enfants, sont exposées, vendues, ou encore imprimées sur des accessoires (Applegate et Grupper, 2013 ; Matsuzawa, 2017), sans que leur auteurisation soit remise en question. Ainsi, il est possible de posséder un foulard aux motifs dessinés par la femelle chimpanzé Aï ou un tableau peint par le mâle chimpanzé Congo dans les années 1960. Mais alors si un chimpanzé est reconnu comme auteur de son dessin, pourquoi ne serait-ce pas le cas pour un enfant en bas âge qui gribouille, un ver de terre qui se traîne dans de la peinture, ou une machine programmée pour créer ? Quelle définition pour la création dès lors qu'elle est précoce, non humaine, voire non biologique ? La notion même d'intelligence semble

finalement être interrogée ici : les capacités cognitives sous-jacentes à l'intention de la réalisation d'une part, celles nécessaires à sa reconnaissance de la réalisation d'autre part. De cette notion d'intelligence « créatrice » émanent d'autres concepts, éthiques et juridiques, à l'origine de la personnalité juridique. Cet article abordera ces différentes notions afin de discuter de l'auteurisation (c'est-à-dire la caractérisation d'un auteur à part entière et sa reconnaissance) ainsi que de l'appropriation d'une création, plus particulièrement d'un dessin ou d'une peinture, réalisée par un animal ou une machine.

De l'intention de créer...

En philosophie, l'éthique de la vertu se concentre sur le caractère et les motivations de l'individu-agent, tandis que le conséquentialisme se concentre sur les conséquences des actions engagées, sans considération des moyens d'y parvenir ni de l'intentionnalité potentielle. Ainsi dans le cadre du dessin, l'éthique de la vertu permet d'envisager l'art comme le seul processus intentionnel, sans prendre en compte ce que sera le résultat final tandis que le conséquentialisme, au contraire, envisagera l'art en tant que seul produit fini, sans tenir compte du processus de réalisation. L'intentionnalité d'un animal dessinateur apparaît donc comme un point clé pour définir si ce dernier est véritablement auteur de son dessin. Pour certains auteurs, l'intentionnalité va même jusqu'à caractériser l'art et l'artiste (Beardsley, 1970). Pour d'autres, l'intentionnalité se définit par cinq éléments : (1) un désir de résultat ; (2) une croyance quant à l'action menant à ce résultat ; (3) une intention d'exécuter l'action ; (4) la conscience de l'accomplissement de l'intention tout en exécutant l'action ; enfin (5) l'habileté à exécuter l'action (Malle et Knobe, 1997). En éthologie cognitive, des études ont montré que des individus d'espèces animales telles que les pigeons, les rats ou les primates non humains sont capables d'anticiper

leurs actes (Le Neindre *et al.*, 2018). Les individus de la plupart de ces espèces sont également conscients de leurs prises de décisions et des conséquences de ces dernières (Le Neindre *et al.*, 2018). Ces degrés d'anticipation de l'action rappellent les éléments (1) et (2) définis par Malle et Knobe (1997), tandis que le degré de conscience de ces animaux se rattache aux éléments (3) et (4) qui définissent l'intentionnalité, toujours selon ces mêmes auteurs (1997). En appliquant ces deux capacités (anticipation et conscience) au comportement de dessin, le conséquentialisme qualifierait le singe ET le ver comme auteurs de leur dessin tandis que l'éthique de la vertu considérerait le singe comme seul auteur, conscient de ses actions.

Néanmoins, il est nécessaire de déterminer si ces deux capacités sont réellement sous-jacentes au dessin animal. En effet, certains chercheurs critiquent le fait que les dessins d'animaux ne soient qu'induits par les expérimentateurs qui les invitent à dessiner, voire les y incitent (Tomasello et Call, 2004). D'ailleurs, dans la nature, aucun comportement spontané de dessin n'a été reporté à ce jour, ni chez les éléphants, ni chez les grands singes, mais il est commun pour des hominidés captifs de manipuler crayons et pinceaux sur des feuilles de papier, voire de dessiner au doigt sur tablettes tactiles (Martinet et Pelé, 2020). Ainsi, certains chimpanzés maintiennent leur activité graphique en dehors de tout renforcement alimentaire (*fig. 1*), ce qui indique l'intérêt qu'ils portent à l'action. Au-delà de la sensation directement liée au mouvement locomoteur, le retour visuel aurait aussi une propriété de renforcement, c'est-à-dire que la vue du dessin incite l'animal à dessiner encore. En effet, le comportement de dessin sur écran tactile diminue lorsque le tracé réalisé par le chimpanzé devient invisible (Tanaka *et al.*, 2003). Les différentes études chez les chimpanzés permettent donc de soutenir l'argument

selon lequel le seul fait de dessiner aurait une propriété de renforcement chez ces animaux. Quoique le comportement de dessin ne soit pas spontané chez les chimpanzés, le simple acte de marquer ainsi que les propriétés de ces marques modifient les futures actions des sujets. Thierry Lenain emploie d'ailleurs le terme de « marques perturbatrices » pour qualifier les créations picturales des primates (Lenain, 1990). Par contre, si le ver de terre ne laisse pas de trace derrière lui, cette absence de trace n'impactera pas ses déplacements sur la feuille contrairement à l'hominidé. Schiller (1951) alla plus loin et proposa à une jeune femelle chimpanzé, nommée Alpha, des feuilles blanches présentant des figures géométriques. Cette dernière marqua les feuilles différemment en fonction des stimuli présentés, faisant ainsi émerger, pour la première fois, la question de l'intentionnalité derrière ces « gribouillages » d'animaux, d'apparence si « simplistes ». Cependant, bien qu'un nombre important d'études fût réalisé chez les grands singes, et les chimpanzés en particulier, aucun dessin représentatif ne fut jamais constaté, et les chercheurs comparent généralement leurs productions à celles de jeunes enfants humains (Martinet et Pelé, 2020). Grâce à l'utilisation d'indices mathématiques fractals évaluant la représentativité d'un dessin, Martinet *et al.* (2021) montrèrent, néanmoins, que des dessins de chimpanzés, même s'ils ne sont pas aussi efficaces que ceux d'enfants, ne sont pas pour autant aléatoires.



Figure 1. *Le chimpanzé Ai dessinant sur une feuille de papier.*
© Tetsuro Matsuzawa, Primate Research Institute, Kyoto University

Ces éléments laissent à penser que certains individus de certaines espèces - en particulier les grands singes - semblent interagir avec leur dessin, dans le sens où ce qui est créé influence ce qui le sera. Le concept d'agentivité (Gell, 1998 ; Blattner et al., 2020) peut donc être attribué à ces animaux d'un point de vue psychologique, par exemple comme acteur du monde agissant sur son environnement ; mais aussi d'un point de vue philosophique, si nous considérons les grands singes comme étant capables de se reconnaître en tant qu'individus et d'avoir un certain sens de la moralité, capacités qui définissent l'essence même d'une personne. En effet, longtemps réservé aux seuls êtres humains, le terme « personne » réfère à « un être intelligent pensant, qui a raison et réflexion, et qui peut se considérer soi-même

comme soi-même, la même chose pensante en des temps et des lieux différents » (Locke, 1975).

... à la reconnaissance de l'auteur

Il n'existe pour l'heure aucune recherche sur les capacités des grands singes à s'approprier leur création. Cependant, conscience et moralité apparaissent comme deux concepts essentiels au processus d'auteurisation permettant - ni plus ni moins - à l'individu de devenir une « personne », puis d'acquérir des droits par sa personnalité juridique. La reconnaissance de soi et la moralité sont donc importantes en ce qui concerne l'auteurisation et l'appropriation d'une création, puisqu'elles ne sont pour le moment attribuées qu'aux seuls êtres humains « pourvus d'une âme et d'une conscience ».

Malebranche en son temps affirmait que « dans les animaux il n'y a ni intelligence, ni âme ». Par la suite, Marx a développé dans les *Thèses sur Feuerbach* : « On peut distinguer les hommes des animaux par la conscience, par la religion et par tout ce que l'on voudra. » Conscience, moralité et reconnaissance de soi séparent donc les animaux des hommes et leur renieraient tout droit, dont celui d'être auteurs d'une création - puisque par définition seules les personnes peuvent être auteurs. Pourtant, remettant en question cette scission hypothétique, de nombreux chercheurs - depuis Charles Darwin qui attribua une conscience aux individus d'espèces sociales - se penchent sur l'existence de cet intellect chez les animaux. Donald Griffin (2013) suggère que c'est au sein des processus de communication - qui impliquent dialogues et négociations - qu'il convient d'explorer les comportements intentionnels et les processus qui seraient contrôlés par la conscience. Les recherches effectuées sur les grands singes capables d'utiliser le langage des signes

ou des symboles ont montré qu'ils pouvaient parler des autres et d'eux-mêmes comme d'entités distinctes et autonomes. La conscience du soi autobiographique (Le Neindre *et al.*, 2018) s'est également avérée présente chez de nombreuses espèces animales (primates, cétacés, oiseaux) par l'utilisation du test du miroir qui montre que les sujets sont capables de s'identifier et de s'y reconnaître dans ce dernier (Gallup Jr *et al.*, 2002). De même, certains animaux sociaux montrent aussi une théorie de l'esprit, c'est-à-dire une capacité à imputer un soi mental à des congénères ou à comprendre ce qu'ils regardent, ce qu'ils ont l'intention de faire, voire même de connaître leurs croyances. Enfin, des observations de chimpanzés et d'éléphants, mais aussi des expériences menées avec des rats, ont montré que ces animaux possèdent un certain degré d'empathie et de moralité (De Waal, 2006). De ces nouvelles découvertes, la qualification de « personne » pourrait donc être attribuée aux animaux capables de se reconnaître et de faire preuve de moralité. Et ces derniers, par l'attribution d'une personnalité juridique, pourraient se voir accorder la propriété de leur création et être reconnus comme de véritables auteurs.

La personnalité juridique

La personnalité juridique permet à l'individu qui la possède d'être titulaire de droits et de devoirs. De nombreux philosophes défendent la non-reconnaissance des animaux comme personnes juridiques, argumentant leur absence de conscience et de raison. Une telle dialectique a vite été défaits, en premier lieu par l'exemple des cas marginaux soulevé par Peter Singer (DeGrazia, 1990). En effet, de nombreux humains - les enfants en bas âge, des adultes fortement handicapés mentalement ou dans le coma - ne possèdent pas les capacités susnommées, mais possèdent pourtant bien une personnalité juridique. Le raisonnement de Peter Singer est de dire que si ces individus possèdent une personnalité juridique, alors pourquoi ne pas attribuer

cette dernière à des animaux qui auraient des capacités égales voire supérieures à ces individus humains. Le jeune enfant ou l'adulte handicapé qui griffonne a des droits et sera reconnu comme auteur de son dessin, quels que soient ses niveaux d'intentionnalité et de conscience de son dessin. Alors pourquoi ces droits ne seraient-ils pas attribués aux animaux qui possèdent tout au moins des capacités similaires à ces êtres humains? La question est complexe et plusieurs actions en faveur de l'attribution de droits envers les animaux ont déjà été initiées par des associations de protection animale (Wise, 2010).

En 2016, en Indonésie, un macaque noir sauvage (et prénommé par la suite Naruto) se prit en photo avec l'appareil d'un photographe professionnel. Ce « selfie » fit le tour du monde et fut très vite considéré comme une aubaine financière par le propriétaire de l'appareil photo qui en réclama les droits d'auteurs. L'association PÉTA porta plainte et entama un procès contre le propriétaire de l'appareil photo afin que le seul singe soit reconnu comme auteur du cliché et que les droits d'auteur lui soient versés si cette image venait à être utilisée à des fins commerciales. Cependant, le vide juridique concernant le droit d'auteur non humain et le fait que Naruto ne soit pas reconnu comme une personne juridique mena le juge à rejeter la demande de PÉTA (Rosati, 2017). Ainsi, reconnaître les animaux possédant une conscience comme personnes juridiques apparaît comme un prérequis pour qu'ils soient véritablement et pleinement auteurs de leur dessin. En 2019, la déclaration de Toulon reprend la déclaration de Cambridge (2012) qui stipule que la plupart des animaux ayant les substrats neurologiques de la conscience devraient posséder une personnalité juridique (Regad et Riot, 2019). La théorie des droits des animaux est de plus en plus discutée et formalisée, permettant aux animaux, selon leur espèce, leur agentivité et leurs interactions avec les êtres humains d'avoir des droits

reconnus (Donaldson et Kymlicka, 2011). L'auteurisation d'une création picturale animale s'intègre parfaitement dans cette théorie.

Et l'intelligence artificielle ?

En 2018, un tableau réalisé par une intelligence artificielle s'est vendu aux enchères pour la somme de 432 500 \$ (Doherty, 2019). Une formule mathématique fait guise de signature, mais l'algorithme à la base de cette création a été développé par le collectif de chercheurs *Obvious*¹¹. Une intelligence artificielle pourrait-elle être reconnue comme auteure de ses créations ? Cet algorithme pourrait-il recevoir des droits d'auteurs ? De même, il existe des algorithmes simples d'accès permettant d'appliquer le style d'un peintre choisi à une image donnée, comme la plateforme Playform par Artrendex Inc. (fig. 2). Dans un tel cas, qui est vraiment l'auteur de l'image finale ? L'intelligence artificielle elle-même ? Les informaticiens ayant programmé l'algorithme ? Les artistes des œuvres ayant permis l'entraînement de l'algorithme ? La démarche scientifique et l'argumentation appliquées à l'intelligence animale pourraient certainement s'appliquer à l'intelligence artificielle (Nguyen, 2019). Dans le cas de la création picturale *Obvious*, l'algorithme produit n'est pas intelligent au sens où nous l'avons décrit ci-dessus. Il est à noter que le modèle ayant permis de créer ce tableau a été entraîné sur des peintures existantes, ce qui peut être assimilable au processus d'apprentissage présent chez les êtres humains. Capable de produire un acte précis, ce modèle n'est cependant pas apte à résoudre n'importe quel problème, n'a pas d'actes intentionnels, ni n'est conscient de lui-même. Par contre, certains robots sont capables de se reconnaître dans un miroir (Hart et Scassellati, 2012). De nouveaux champs de recherche sont ainsi créés et développés afin de mieux comprendre le comportement des machines

¹¹ <https://obvious-art.com/>

et leurs capacités émergentes (Rahwan *et al.*, 2019). Des tests spécifiques tels que le test de Turing¹² existent afin de connaître précisément les capacités des machines, notamment en lien avec la création artistique (Bishop et Boden, 2010). Néanmoins, ces tests sont souvent critiqués, car conceptualisés par l'humain avec tous les biais anthropomorphiques que cela présuppose (Sueur et Pelé, 2017 ; Sueur *et al.*, 2020). Or, même si d'aucuns ne souhaitent pas reconnaître les machines comme des artistes, il est important de noter la difficulté pour un être humain de discerner des créations informatiques de créations humaines (Mikalonytė et Kneer, 2021). Par contre, la différence essentielle entre la machine et l'animal apparaît être l'essentialisme. En philosophie, l'essentialisme suppose que l'essence d'une chose précède son existence. La machine a été créée par l'homme et de ce fait, la machine reste (pour le moment) dépourvue de tous ses droits, et ce, au profit de son créateur. Cet aspect philosophique est bien sûr un point de vue défendu par certains, décrié par d'autres, et dépend de nombreux facteurs tels que la nature des agents impliqués (robots articulés ou simples ordinateurs), du degré d'intentionnalité de la création (intentionnelle ou accidentelle) et de la création graphique elle-même (abstraite ou représentative).

¹² Du nom de son créateur, ce test permet d'évaluer une intelligence artificielle en mesurant sa faculté à imiter une conversation humaine.

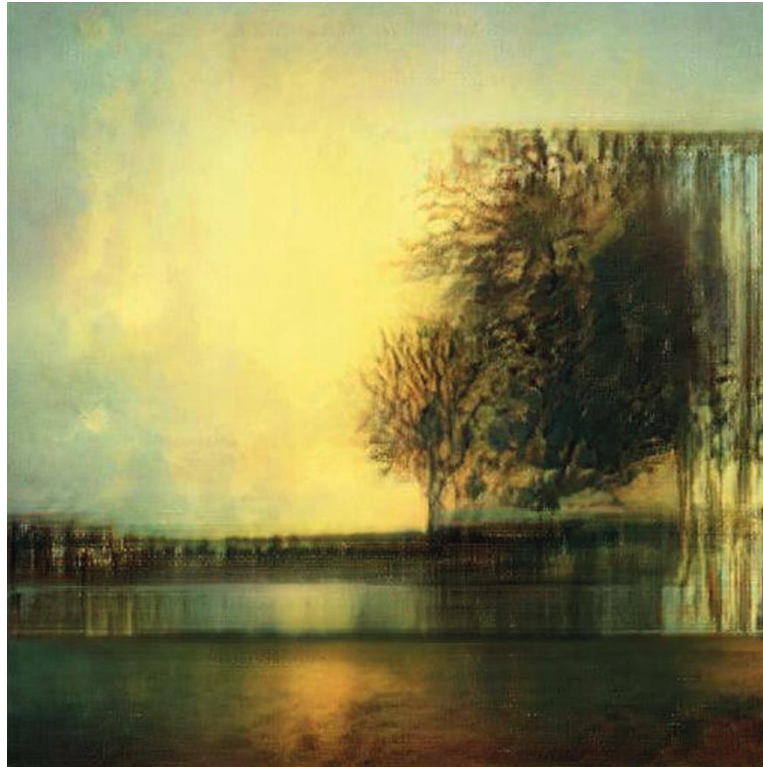


Figure 2. Image tirée d'une photographie et modifiée selon un algorithme d'intelligence artificielle imitant le style de William Turner. La plateforme Playform par Artrendex Inc. a été utilisée.

Conclusion

Les concepts liés à l'auteurisation et l'appropriation d'une création, notamment picturale, sont ceux qui définissent une personne : une entité consciente ayant des droits. La démarche scientifique pour l'application de droits d'auteurs se divise en différentes étapes permettant d'évaluer l'intentionnalité d'un acte et la conscience que l'on a de ce dernier. Ainsi, les machines ne sauraient être reconnues, tout du moins pour le moment, comme auteures de leurs créations. Mais cette conclusion apparaît différente concernant les animaux pourvus de conscience ou pour lesquels nous sommes aujourd'hui en mesure de l'identifier.

Chez les grands singes, même si seuls des individus captifs prennent plaisir à dessiner, cette enculturation (Tomasello et Call, 2004) ne doit pas empêcher de leur reconnaître l'auteurisation et l'appropriation de leurs créations, puisqu'un processus

similaire d'apprentissage est observé chez l'homme. En effet, dessiner n'est pas plus spontané chez *Homo sapiens* que chez les autres hominidés et c'est un long processus d'observation et d'acquisition du geste qui attend les jeunes enfants. Par simple volonté de jouer, ces derniers peuvent produire des dessins (figuratifs ou non) sans intention de représentation, ni de reconnaissance de leur acte. Il n'empêche que ces jeunes enfants sont reconnus comme auteurs de leur création. Alors pourquoi pas les animaux? Certains primatologues incluent déjà les primates avec lesquels ils travaillent dans la liste d'auteurs de leurs travaux, publications ou productions (Savage-Rumbaugh et al., 2007 ; Applegate et Grupper, 2013 ; Matsuzawa, 2017). Mais prendre la décision d'auteuriser les animaux à la place de ces derniers, ne serait-ce pas finalement leur confisquer leur agentivité (Blattner et al., 2020) ? Pour être auteur d'une création, doit-on exprimer une quelconque volonté de l'être ? Une chose est sûre, l'auteurisation continue de diviser la communauté, mais nous engage à penser toujours plus loin notre relation aux non-humains.

Bibliographie

- Applegate K. & Grupper A. 2013. *The one and only Ivan* [livre audio]. New York, HarperAudio.
- Beardsley M.C. 1970. « The aesthetic point of view ». *Metaphilosophy*, 1(1), p. 39-58.
- Bishop M. & Boden M.A. 2010. « The Turing test and artistic creativity ». *Kybernetes*, 39(3), p. 409-413.
- Blattner C.E. *et al.* 2020. « Animal Agency in Community ». *Politics and Animals*, 6, p. 1-22.
- De Waal F. 2006. « Morally Evolved: Primate Social Instincts, Human Morality, and the Rise and Fall of “Veneer Theory” ». In: De Wall F. *et al.*, *Primates and philosophers: How morality evolved*. Princeton, Princeton University Press, p. 1-80.
- DeGrazia D. 1990. « On Singer: More argument, less prescriptivism ». *Behavioral and Brain Sciences*, 13(1).
- Diamond J. 1986. « Animal art: variation in bower decorating style among male bowerbirds *Amblyornis inornatus* ». *Proceedings of the National Academy of Sciences*, 83(9), p. 3042-3046.
- Doherty S.J. 2019. « ART in the Age of Artificial Intelligence ». *Esse*, 97, p. 30-41.
- Donaldson S. & Kymlicka W. 2011. *Zoopolis: A political theory of animal rights*, Oxford, Oxford University Press.
- Gallup Jr G.G. *et al.* 2002. « The mirror test ». In : Bekoff M., Allen C., & Burghardt G. M. (Éds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition*. Chicago, University of Chicago Press, p. 325-333.
- Gell A. 1998. *Art and agency: an anthropological theory*. Oxford, Clarendon Press.
- Griffin D.R. 2013. *Animal minds: Beyond cognition to consciousness*. Chicago, University of Chicago Press.
- Hart J. & Scassellati B. 2012. « Mirror perspective-taking with a humanoid robot », dans. *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 26, No. 1, pp. 1990-1996)
- Kawase H. *et al.* 2013. « Role of Huge Geometric Circular Structures in the Reproduction of a Marine Pufferfish ». *Scientific Reports*, 3(1), p. 2106.
- Lenain T. 1990. *La peinture des singes*. Paris, Éditions Alternatives.
- Le Neindre P. *et al.* 2018. *La conscience des animaux*. Versailles, Éditions Quæ.
- Locke J. 1975. « Of Identity and Diversity ». In : *Essay Concerning Human Understanding*. Berkeley, University of California Press.
- Malle B.F. & Knobe J. 1997. « The folk concept of intentionality ». *Journal of experimental social psychology*, 33(2), p. 101-121.
- Martinet L. *et al.* 2021. « New indices to characterize drawing behavior in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) ». *Scientific Reports*, 11(1), p. 3860.
- Martinet L. & Pelé M. 2021. « Drawing in nonhuman primates: What we know and what remains to be investigated. ». *Journal of Comparative Psychology*, 135(2), p. 176-184 (DOI : <https://doi.org/10.1037/com0000251>).
- Matsuzawa T. 2017. « The 40th anniversary of the Ai Project: the commemorative gift is a silk scarf painted by Ai the chimpanzee ». *Primates*, 58, p. 261-265.
- Mikalonytė E.S. & Kneer M. 2021. « Can Artificial Intelligence Make Art? ». *arXiv : 2104.07598 [cs]*. Available at SSRN: <https://ssrn.com/abstract=3827314> or <http://dx.doi.org/10.2139/ssrn.3827314>

- Nguyen P. 2019. « The Monkey Selfie, Artificial Intelligence and Authorship in Copyright: The Limits of Human Rights ». *Pub. Int. LJNZ*, 6, p. 121.
- Rahwan I. *et al.* 2019. « Machine behaviour ». *Nature*, 568, 7753, p. 477-486.
- Regad C. & Riot C. 2019. « La Déclaration de Toulon du 29 mars 2019 sur la personnalité juridique de l'animal : pourquoi? quelle potentialité juridique? », dans. In *Démocratie et Justice*
- Rosati E. 2017. « The Monkey Selfie case and the concept of authorship: an EU perspective ». *Journal of Intellectual Property Law & Practice*, 12(12), p. 973-977.
- Savage-Rumbaugh S. *et al.* 2007. « Welfare of Apes in Captive Environments: Comments On, and By, a Specific Group of Apes ». *Journal of Applied Animal Welfare Science*, 10(1), p. 7-19.
- Schiller P.H. 1951. « Figural preferences in the drawings of a chimpanzee. ». *Journal of Comparative and Physiological Psychology*, 44(2), p. 101.
- Sueur, C., Forin-Wiart, M. A., & Pelé, M. (2020). Are they really trying to save their buddy? the anthropomorphism of animal epimeletic behaviours. *Animals*, 10(12), 2323.
- Sueur C. & Pelé M. 2017. « Editorial : Anthropomorphism, between merits and demerits ». *Lettre du CEERE, Centre Européen d'Enseignement et de Recherche en Ethique*, 107, p. 1.
- Tanaka M. *et al.* 2003. « Finger drawing by infant chimpanzees (Pan troglodytes) ». *Animal Cognition*, 6(4), p. 245-251.
- Tomasello M. & Call J., 2004, « The role of humans in the cognitive development of apes revisited ». *Animal cognition*, 7(4), p. 213-215.
- Vancatova M., 2008, « Creativity and innovative behaviour in primates on the example of picture-making activity of apes ».
- Watanabe S. 2012. « Animal Aesthetics from the Perspective of Comparative Cognition », dans Watanabe S. & Kuczaj S. (éd.), *Emotions of Animals and Humans*. Springer Japan, coll. « The Science of the Mind », p. 129-162.
- Wise S.M. 2010. « Legal personhood and the nonhuman rights project ». *Animal L.*, 17, p. 1.
- Zagala S. 2004. « Vanuatu sand drawing ». *Museum International*, 56, 1-2, p. 32-35.

Chapitre 5

Discussion



Que ce soit chez l'être humain ou chez d'autres primates, les différentes études menées dans ce projet de thèse ont démontré la capacité du DL à discriminer des dessins selon différents facteurs. Une des principales difficultés rencontrées dans ces études était l'interprétabilité de ces résultats : bien qu'un CNN soit capable de classer dans une certaine mesure ces dessins, comment l'expliquer ? Quelles variables le modèle a-t-il découvertes et utilisées ? Afin de répondre à ces questions, nous avons utilisé différentes techniques manipulant les couches de convolution et les prédictions du modèle. Cependant, bien qu'informatives, ces techniques ne permettent pas une compréhension totale des caractéristiques considérées par le modèle.

Avant de s'être intéressé au deep learning, ce projet de recherche s'est concentré sur la définition de nouveaux indices mathématiques permettant d'analyser le comportement de dessin de manière plus objective. Les limites et perspectives de ces indices ne seront pas discutées ici, étant donné que ce travail a été réalisé dans le cadre du doctorat de Lison Martinet (2021), qui a permis d'ouvrir la voie à de nouvelles approches quant à l'étude du comportement de dessin à travers des indices mathématiques et m'a permis de développer mon propre travail de doctorat.

I. Limites et perspectives du deep learning

Ces dernières années, les modèles de DL se sont énormément démocratisés et permettent aujourd'hui d'obtenir des performances remarquables dans différents domaines tels que la reconnaissance d'image, le traitement automatisé du langage, ou encore la reconnaissance vocale. Seulement, ces modèles reposent sur des millions (voire des milliards) de paramètres et des architectures qui peuvent varier selon la tâche, rendant ainsi complexe l'interprétation de ces derniers. Pour ces raisons, les modèles de DL sont généralement considérés comme des boîtes noires et

comprendre les décisions prises par de tels modèles en termes de classification reste un des principaux défis du DL à ce jour (Buhrmester et al., 2021 ; Castelvechi, 2016 ; Lei et al., 2018). Comment exploiter ces décisions afin d'en apprendre plus sur ce que ces modèles ont pu extraire et découvrir des données utilisées ?

1) Que voit un modèle de deep learning ?

L'intérêt d'utiliser le DL pour analyser le dessin est d'avoir une approche plus objective qu'une interprétation faite par un œil humain. Seulement, la « vision » d'un modèle DL est-elle comparable à celle d'un humain ?

a. CNN et système visuel humain

Les réseaux de neurones artificiels sont inspirés du cerveau humain, mais il est légitime de se demander si cette ressemblance ne repose que sur la forme, ou si le fonctionnement de ces derniers est également similaire. Ces questions, étroitement liées à l'interprétation des modèles de DL, ont beaucoup été étudiées ces dernières années. En réalité, de nombreuses similarités entre le système visuel humain et les CNNs existent (Geirhos et al., 2022 ; Kubilius et al., 2016 ; Rajalingham et al., 2018). Un lien existe notamment entre le système visuel humain et les couches des CNNs : les aires visuelles de bas niveau sont mieux décrites par les premières couches de convolution, tandis que les aires visuelles de plus haut niveau sont mieux décrites par des couches de convolution plus profondes (Cadena et al., 2017 ; Cichy et al., 2016 ; Eickenberg et al., 2017 ; Seibert et al., 2016).

Dans les lois de la Gestalt, la loi de clôture affirme que notre système visuel aura tendance à compléter des figures inachevées. Cette loi et d'autres phénomènes comme l'effet Thatcher ou la loi de Weber sont présents dans des réseaux neuronaux entraînés sur des images du monde réel, mais ne le sont pas dans des réseaux composés de poids aléatoires (Jacob et al., 2021 ; Kim et al., 2021). Ainsi, il est

fondamental d'être conscient que le degré de similarité entre le cerveau humain et le modèle utilisé n'est pas inhérent au DL, mais dépendra des poids, et donc de l'entraînement du modèle. En utilisant un CNN où le phénomène étudié est présent et en distinguant préalablement les neurones responsables de son activation, il serait alors possible de quantifier la présence de ce phénomène sur différents dessins. Une telle approche permettrait d'étudier et quantifier l'émergence de certains phénomènes, comme ceux cités précédemment. Cependant, cette approche reste limitée, puisque certains phénomènes ne semblent pas présents dans les CNNs, comme l'occlusion (Jacob et al., 2021).

Si les CNNs et la vision humaine présentent des points communs, il reste facile de se convaincre qu'un écart important les sépare (Xu & Vaziri-Pashkam, 2021). En partant par exemple d'une photo de panda, il est possible d'altérer la valeur de chaque pixel pour que l'image, perçue par un humain, ressemble toujours sans hésiter à un panda, mais qu'un modèle de DL le classifie à 99 % comme un gibbon (Goodfellow et al., 2015). Un tel exemple est appelé *adversarial example*. Malgré des différences, les CNNs décrivent relativement bien le système visuel humain ainsi que l'activité des aires visuelles comparés aux autres modèles généralement utilisés (Cadieu et al., 2014 ; Cichy et al., 2016 ; Yamins et al., 2014).

b. Texture et forme

Savoir si la perception des CNNs est comparable à celle des humains est important, mais ce n'est pas la seule question à se poser. Un aspect fondamental de l'analyse du comportement de dessin est la forme de ces productions. Mais la forme est-elle réellement ce sur quoi se basent les CNNs pour la classification ?

Geirhos et al. (2018) ont démontré qu'une architecture entraînée sur ImageNet sera sujette à un biais de texture¹³ (le modèle aura tendance à beaucoup plus utiliser la texture que la forme), et qu'en utilisant un autre jeu d'entraînement, Stylized-ImageNet (une version stylisée d'ImageNet), cette même architecture apprendrait des représentations bien plus liées à la forme, dont l'information est encodée dans les couches de convolution de haut niveau (Islam et al., 2021). Cette question de texture et de forme est fondamentale pour le dessin. En effet, la texture sera dépendante des instruments utilisés plutôt que du dessinateur lui-même. Ainsi, pour utiliser de manière optimale les modèles de DL sur des dessins, il est premièrement important d'avoir un protocole de récolte de données permettant d'avoir des dessins de texture homogène. Si l'instrument pour certains dessins est une craie et pour d'autres un feutre, ou que le support est du papier pour certains dessins et de la toile pour d'autres, le réseau de neurones convolutif utilisé risquera de bien plus considérer cet aspect de texture plutôt que les formes représentées. Pour ces raisons, les tablettes tactiles présentent un intérêt crucial : la texture de chaque dessin sera similaire, étant donné que chaque pixel du support pourra être exactement de la même couleur. De plus, bien que les modèles des études de ce projet de thèse aient été entraînés de nouveau à travers une étape de fine-tuning, les modèles sont à la base pré-entraînés sur ImageNet, comme de nombreuses architectures. Ainsi, toute l'information liée à la forme n'a peut-être pas été utilisée. Il sera donc nécessaire à l'avenir de prêter une plus grande importance au

¹³ Il est à noter que ce que signifie la « texture » est complexe à définir, car elle semblerait selon Gatys et al. (2015) intrinsèquement liée au « style » d'une image, comme considéré dans l'Article 5, mais, dans les études opposant texture et forme, si une image représente par exemple un éléphant, la « texture » de l'éléphant sera sa peau, ce qui ne correspondrait pas réellement à ce que nous pourrions appeler le « style » d'une image.

jeu d'entraînement utilisé de base afin de s'assurer qu'il permette d'extraire le maximum d'informations liées à la forme.

2) Interprétabilité du deep learning

Bien que les modèles de DL soient capables d'apprendre des représentations d'objets, d'animaux, ou encore de scènes, et qu'ils présentent des points communs avec le système visuel humain, cela ne garantit pas que les prédictions de ces modèles soient basées sur ce que nous voyons. Pour illustrer cela, un exemple très connu de la littérature est un modèle de classification de photos de loups et de huskies. Le modèle en question a été volontairement entraîné sur un jeu de données où l'arrière-plan de chaque photo de loup était de la neige, ce qui n'était pas le cas pour les photos de huskies. Bien que ce modèle soit très performant, il ne se basait en réalité pas sur l'animal présent sur la photo, mais simplement sur la présence ou non de neige (Ribeiro et al., 2016). Pour cette raison, il est primordial d'avoir des outils permettant de comprendre plus en détail les décisions prises par le modèle. L'interprétation du deep learning peut être étudiée à deux niveaux. Premièrement, d'un point de vue local, chaque image peut être analysée afin de comprendre les régions qui ont mené le modèle à sa décision. D'un point de vue global, au lieu d'analyser une image en particulier, il est possible d'analyser le comportement du modèle.

a. Interprétation visuelle de réseaux neuronaux

À partir d'un CNN, différentes explications peuvent être données. Dans l'Article 5 par exemple, l'interprétation proposée repose sur l'extraction des activations de chaque couche de convolution, qui représentent des caractéristiques de différents niveaux. Dans l'Article 4, l'explication proposée est liée à la comparaison entre les prédictions de différents modèles. En plus de ces prédictions, un algorithme

permettant de visualiser les cartes de chaleur, grad-CAM, a été utilisé (Selvaraju et al., 2017). Dans de nombreux cas, les cartes de chaleur ne semblaient pas très informatives, se concentrant parfois sur des caractéristiques comme le visage représenté sur le dessin, parfois sur des parties totalement vides du dessin, comme les bords du support. Il est donc difficile de savoir si de telles cartes de chaleur sont interprétables. Comment alors quantifier l'interprétabilité d'un modèle ? Theodorus et al. (2020) ont proposé un score d'interprétabilité des cartes de chaleur, spécifiquement pour le dessin, basé sur la correspondance entre les traits du dessin et la carte de chaleur de ce dernier. Ils ont ainsi démontré que l'architecture de BagNet (Brendel & Bethge, 2019) permettait d'avoir des cartes de chaleur plus pertinentes sur les dessins que d'autres architectures, comme celle de ResNet (He et al., 2016). Il serait donc alors pertinent d'utiliser une telle architecture afin d'espérer découvrir de nouvelles caractéristiques du dessin. En plus de cette méthode de visualisation, de nombreuses autres méthodes existent et pourraient également être appliquées au dessin (Mahendran & Vedaldi, 2015 ; Springenberg, 2016).

b. Détection de concept

L'approche expliquée précédemment vise à découvrir les caractéristiques, les concepts d'intérêt, pour la classification. Mais l'inverse est également possible : voir comment un concept donné impacte la classification. Cette méthode, nommée Testing with Concept Activation Vectors (TCAV), permet de générer des explications globales de réseaux neuronaux (Kim et al., 2018). En considérant par exemple un modèle classifiant des photos de zèbres, cette méthode permet de répondre à la question suivante : comment le concept de rayure influence la classification ? Pour répondre à cette question, un jeu de données représentant le concept de rayures

doit être utilisé. Contrairement aux cartes de chaleur, TCAV n'analyse pas une image en particulier, mais le modèle dans son entièreté. Appliquée au dessin, cette méthode permettrait par exemple, en utilisant le concept de « visage », de comparer le degré d'anthropomorphisme de dessins représentant des animaux, et pouvoir ainsi comparer l'anthropomorphisation de ces animaux selon la culture du dessinateur. Cette méthode présente cependant des limites. Premièrement, des concepts très abstraits, comme l'amour ou encore le bonheur seront difficilement compris par le modèle utilisé (Molnar, 2020). Ainsi, le concept considéré doit être clairement défini et assimilable par le modèle. Une autre limite, toujours liée à l'intégration du concept d'intérêt, est la nécessité d'un jeu de données représentant ce concept. Ainsi, afin d'étudier l'anthropomorphisation de dessins d'animaux, il faudrait un jeu de données de dessins de visage, ce qui nécessiterait donc un travail supplémentaire. Pour cela, la base de données de Google Quick, Draw !¹⁴, contenant plusieurs millions de dessins, peut permettre d'extraire différents concepts (comme le concept de visage).

c. Approche non supervisée

En étudiant le comportement de dessin avec le DL, la question que nous nous posons est la suivante : quelles différences existe-t-il entre ces catégories ? Cependant, dans la grande majorité des cas, les modèles de DL sont utilisés pour leur efficacité, et l'objectif est d'avoir le modèle le plus performant. Ainsi, habituellement, l'interprétabilité a pour but de répondre aux questions suivantes : comment améliorer ce modèle ? Quels biais et erreurs y sont présents ? Le but du DL est en général d'automatiser des tâches réalisables par des humains. Dans le cadre du comportement de dessin, l'objectif est totalement différent : le but est de découvrir

¹⁴ <https://quickdraw.withgoogle.com/>

ce qui permet de différencier les catégories. Pour cette raison, d'autres approches, comme le clustering, devraient être beaucoup plus explorées. Combiner des réseaux neuronaux et un algorithme de réduction de dimension comme l'algorithme t-SNE (Van der Maaten & Hinton, 2008), très utilisé en deep learning (Liu et al., 2019 ; Perez & Tah, 2020 ; Svantesson et al., 2023) permettrait de visualiser les dessins sur un plan et donc voir quels dessins sont les plus similaires. En considérant des dessins d'enfants de différents âges, une telle approche permettrait de voir quels groupes seraient les plus proches. Est-ce que les dessins seraient regroupés selon différentes catégories d'âges ? Quels groupes seraient alors proches ? Y retrouverions-nous les groupes théorisés par Luquet ou Baldy ? Les groupes seraient-ils au contraire plutôt basés sur d'autres aspects du dessinateur, comme sa culture ? De nombreuses autres possibilités pourraient être envisagées. Par exemple, en considérant des dessins produits par des individus de différentes cultures, nous pouvons nous demander si des groupes relatifs à chaque culture seraient observables et d'essayer d'en comprendre les causes, Les distances entre les dessins de chaque culture seraient-elles alors corrélées aux distances entre les différents pays correspondants ?

Cette approche pourrait également être appliquée sur des dessins d'autres primates, comme ceux de l'Article 5. En effet, l'Article 5 utilise une différence saisonnière connue a priori entre les dessins. Cette approche exploratrice pourrait être repensée de manière non supervisée, notamment à travers du clustering. Quels dessins seraient alors les plus proches ? Y aurait-il différents groupes ? Seraient-ils basés sur les saisons, ou encore sur une autre échelle temporelle ? Le clustering ne serait cependant pas nécessairement en mesure de répondre à toutes ces questions, étant donné que les variables à la source de ce clustering n'auront pas obligatoirement été récoltées lors de la prise de données.

3) Des bases de données plus complètes

Les méthodes d'interprétation ne sont pas les seules limites à l'étude du comportement de dessin. Comme expliqué précédemment, le DL est majoritairement utilisé pour ses performances et non son interprétabilité. Cette remarque est également valable pour l'étude de dessins en utilisant du DL. Ainsi, les différents groupes considérés dans l'étude du dessin par le DL sont bien plus axés sur le produit du dessin que le processus ou le dessinateur en lui-même. En d'autres termes, les modèles reposent très fortement sur les différences entre les objets représentés et l'objectif est de les discriminer au mieux, sans nécessairement interpréter ou comprendre les différences entre les groupes, ni l'aspect temporel du dessin.

Ainsi, un gros travail reste à faire sur les bases de données de dessins pour l'analyser d'un point de vue comportemental. Considérer des informations liées à l'environnement du dessinateur permettrait d'en apprendre plus sur ce comportement. En effet, des variables comme la culture, le niveau d'éducation, le niveau social, l'âge, le genre, la motivation, les conditions du dessin, ne sont pas, ou très peu, considérées dans de telles analyses, comme montré dans l'Article 1, mais pourraient permettre d'importantes découvertes liées au comportement de dessin, tant en utilisant de la classification que du clustering.

Une fois le dessin terminé, il peut être intéressant de demander au dessinateur d'annoter sa production. Cela pourrait être fait par exemple à travers de la segmentation ou encore en lui demandant d'encadrer et annoter les parties du dessin d'une boîte englobante (définir un rectangle sur l'image autour d'un motif comme un objet, un visage). Cette approche permettrait une compréhension plus précise des différentes parties du dessin, et plus encore de l'intention du dessinateur sur les

différentes parties de sa production. En utilisant cette méthode pour des dessins d'enfants, le modèle pourrait alors comprendre plus simplement à quelle partie correspond un objet donné. D'autres technologies pourraient également être utilisées en plus du deep learning. À l'aide d'oculométrie (plus connue sous le nom d'eye-tracking), il serait par exemple possible de demander à des personnes de regarder des dessins, afin de voir les parties jugées importantes par ces derniers. Cette approche permettrait notamment de comparer les cartes de chaleur aux régions sur lesquelles les individus observant le dessin se concentrent, pour par exemple juger de l'interprétabilité et la pertinence des cartes de chaleur. Une autre utilisation de l'eye-tracking pourrait être de suivre le regard d'un individu au moment où il dessine, ce qui pourrait permettre de tester le lien entre son regard et son tracé. L'eye-tracking permettrait-il de prédire la future position du crayon ? La qualité de cette prédiction changerait-elle avec l'âge ? De manière plus générale, analyser des gribouillis pourrait être intéressant. En effet, les gribouillis analysés dans l'Article 3, produits par des enfants et des adultes, étaient significativement différents. Malheureusement, le nombre de gribouillis utilisé n'est pas suffisant pour entraîner de manière adéquate un modèle de DL. Une base de données de gribouillis plus conséquente pourrait permettre d'entraîner un modèle de DL afin de classifier des gribouillis d'adultes contre des gribouillis d'enfants. Quelle précision ce modèle atteindrait-il ? Serait-il correct dans la majorité des cas ? Quoi qu'il en soit, de tels dessins resteraient difficiles à classifier d'un œil humain, et ces modèles auraient vraisemblablement une précision supérieure à celle d'un humain.

En plus d'informations liées au dessinateur, l'instrument utilisé pour la récolte des données est également décisif. En effet, en utilisant une tablette tactile, il pourrait être intéressant d'avoir la possibilité d'extraire la pression avec laquelle chaque trait

est produit, qui est notamment liée à la maîtrise des mouvements du dessinateur (Abdulhameed & Rashid, 2022). Certains réseaux de neurones, adaptés aux données temporelles, tels que les réseaux Long short-term memory (LSTM) ou Gated recurrent unit (GRU), permettraient d'analyser l'aspect temporel lié à la pression dont fait preuve l'enfant sur la tablette, pour par exemple estimer le stade de dessin (selon la classification de Luquet par exemple) atteint par ce dernier. De manière plus générale, des modèles de classification de séries temporelles comme les LSTM pourraient permettre la classification de dessins, vus comme processus, et également de traiter chaque trait indépendamment.

II. Conclusion

Durant mon doctorat, j'ai pu assister à différentes conférences sur l'interprétabilité des modèles de machine learning et de deep learning. J'ai ainsi pu constater de la multitude de domaines qui requièrent plus d'explications, et des modèles plus interprétables. J'ai également pu noter que cette notion d'interprétabilité a très majoritairement pour but de diagnostiquer les modèles, dans un souci d'efficacité. Cette approche n'est donc pas en tout point similaire à l'interprétation du comportement de dessin, qui repose plutôt sur la découverte de caractéristiques discriminantes.

L'application du deep learning au comportement de dessin est à ce jour très peu exploré et de nombreux efforts restent à poursuivre, tant sur les bases de données que sur les méthodes d'interprétabilité. En plus de permettre d'en apprendre plus sur le développement cognitif et l'émergence évolutive de ce comportement, des modèles génératifs commencent à voir le jour afin de permettre aux dessinateurs, débutants ou experts, de produire des dessins de meilleure qualité. De manière plus

importante encore, le deep learning pourrait être appliqué en psychologie, sur des dessins d'individus par exemple atteints d'autisme ou de dépression.

Bien que le deep learning offre de nombreuses possibilités quant à l'étude de l'émergence évolutive et du développement du comportement de dessin, les progrès fulgurants réalisés depuis plusieurs années ne cessent de renouveler le domaine. S'il est difficile de savoir quelles techniques d'interprétabilité existeront demain, l'intérêt indispensable et croissant de ce champ de recherche nous laisse imaginer qu'il occupera une place primordiale à l'avenir dans l'étude du comportement de dessin.

Références

- Abdulhameed, S., & Rashid, T. A. (2022). Child Drawing Development Optimization Algorithm Based on Child's Cognitive Development. *Arabian Journal for Science and Engineering*, 47(2), 1337-1351. <https://doi.org/10.1007/s13369-021-05928-6>
- Adi-Japha, E., Levin, I., & Solomon, S. (1998). Emergence of representation in drawing: The relation between kinematic and referential aspects. *Cognitive Development*, 13(1), Article 1.
- Alcopley, L. (1968). Drawings as Structures and Non-structures. *Leonardo*, 3-16.
- Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E., & Fernandez-Anaya, G. (2008). Time-varying Hurst exponent for US stock markets. *Physica A: statistical mechanics and its applications*, 387(24), 6159-6169.
- Ashwin, C. (2016). What is a drawing? *Drawing: Research, Theory, Practice*, 1(2), 197-209.
- Asimov, I. (1942). Runaround. *Astounding science fiction*, 29(1), 94-103.
- Asimov, I. (1967). *Le cycle des robots 1. Les robots*. 5.
- Bailey, G. H. (1982). *Drawing and the drawing activity: A phenomenological investigation* [PhD Thesis].
- Baldy, R. (2005). Dessin et développement cognitif. *Enfance*, 57(1), Article 1.
- Baldy, R. (2011). Fais-moi un beau dessin. *Regarder le dessin de l'enfant, comprendre son évolution*. Paris.
- Bardo, A., Cornette, R., Borel, A., & Pouydebat, E. (2017). Manual function and performance in humans, gorillas, and orangutans during the same tool use task. *American Journal of Physical Anthropology*, 164(4), 821-836. <https://doi.org/10.1002/ajpa.23323>
- Beyries, S., & Joulian, F. (1990). L'utilisation d'outils chez les animaux: Chaînes opératoires et complexité technique. *Paléo*, 2 : 17-26. *The use of tools among animals: chaînes opératoires and technical complexity*) doi, 10.
- Bhatnagar, S., Alexandrova, A., Avin, S., Cave, S., Cheke, L., Crosby, M., Feyereisl, J., Halina, M., Loe, B. S., & Ó hÉigeartaigh, S. (2018). Mapping intelligence: Requirements and possibilities. *Philosophy and theory of artificial intelligence 2017*, 117-135.
- Boesch, C., & Boesch, H. (1990). Tool Use and Tool Making in Wild Chimpanzees. *Folia Primatologica*, 54(1-2), 86-99. <https://doi.org/10.1159/000156428>
- Boysen, S. T., Berntson, G. G., & Prentice, J. (1987). Simian scribbles: A reappraisal of drawing in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, 101(1), Article 1. <http://dx.doi.org/10.1037/0735-7036.101.1.82>
- Brachman, R. J. (2006). Al more than the sum of its parts. *AI Magazine*, 27(4), 19-19.
- Brendel, W., & Bethge, M. (2019). *Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet* (arXiv:1904.00760). arXiv. <https://doi.org/10.48550/arXiv.1904.00760>
- Bresler, L., & Thompson, C. M. (Éds.). (2002). *The Arts in Children's Lives: Context, Culture, and Curriculum*. Springer Netherlands. <https://doi.org/10.1007/0-306-47511-1>

- Brewster, J. M., & Siegel, R. K. (1976). Reinforced drawing in *Macaca mulatta*. *Journal of Human Evolution*, 5(4), 345-347. [https://doi.org/10.1016/0047-2484\(76\)90039-7](https://doi.org/10.1016/0047-2484(76)90039-7)
- Brumm, A., Oktaviana, A. A., Burhan, B., Hakim, B., Lebe, R., Zhao, J., Sulistyarto, P. H., Ririmasse, M., Adhityatama, S., & Sumantri, I. (2021). Oldest cave art found in Sulawesi. *Science Advances*, 7(3), eabd4648.
- Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision : A Survey. *Machine Learning and Knowledge Extraction*, 3(4), Article 4. <https://doi.org/10.3390/make3040048>
- Cachin, C. (1997). *Entropy measures and unconditional security in cryptography* [PhD Thesis]. ETH Zurich.
- Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolia, A. S., Bethge, M., & Ecker, A. S. (2017). *Deep convolutional models improve predictions of macaque V1 responses to natural images* (p. 201764). bioRxiv. <https://doi.org/10.1101/201764>
- Cadiou, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Computational Biology*, 10(12), e1003963. <https://doi.org/10.1371/journal.pcbi.1003963>
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, 538(7623), 20. <https://doi.org/10.1038/538020a>
- Castrignano, A., & Stelluti, M. (1999). Fractal geometry and geostatistics for describing the field variability of soil aggregation. *Journal of agricultural engineering research*, 73(1), 13-18.
- Chappell, P. A., & Steitz, J. A. (1993). Young Children's Human Figure Drawings and Cognitive Development. *Perceptual and Motor Skills*, 76(2), Article 2. <https://doi.org/10.2466/pms.1993.76.2.611>
- Chitiga, M., Owusu-Sekyere, E., & Tsoanamatsie, N. (2014). *Income inequality and limitations of the Gini index : The case of South Africa*.
- Chollet, F. (2020). *L'apprentissage profond avec Python* (machinelearning.fr).
- Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6(1), Article 1. <https://doi.org/10.1038/srep27755>
- Cohn, N. (2012). Explaining 'I can't draw' : Parallels between the structure and development of language and drawing. *Human Development*, 55(4), 167-192.
- Cox, M. (1992). *Children's Drawings* (First Edition). Penguin Books.
- Danic, I., Delalande, J., & Rayou, P. (2006). Enquêter auprès d'enfants et de jeunes : Objets, méthodes et terrains de recherche en sciences sociales. *Revue française de pédagogie. Recherches en éducation*, 156, Article 156. <https://doi.org/10.4000/rfp.653>
- De Araujo, D. B., Tedeschi, W., Santos, A. C. dos, Elias Jr, J., Neves, U. P., & Baffa, O. (2003). Shannon entropy applied to the analysis of event-related fMRI time series. *NeuroImage*, 20(1), 311-317.
- Dechter, R. (1986). *Learning While Searching in Constraint-Satisfaction-Problems*.
- de Condillac, E. B. (1754). *Traité des sensations, à Madame la comtesse de Vassé* (Vol. 1). & se vend a Paris, chez de Bure l'aîné, Quay des Augustins, à Saint Paul.
- DeLoache, J. S. (1991). Symbolic functioning in very young children : Understanding of pictures and models. *Child development*, 62(4), 736-752.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet : A large-scale hierarchical image database* | *IEEE Conference Publication* | *IEEE Xplore*. <https://ieeexplore.ieee.org/abstract/document/5206848>
- Devetak, I., & Winter, A. (2003). Classical data compression with quantum side information. *Physical Review A*, *68*(4), 042301.
- Dexter, E. (2005). *Vitamin D : New perspectives in drawing*. Çin: *Phaidon Yayınları*.
- Ding, L., Luo, Y., Lin, Y., & Huang, Y. (2021). Optimal setting for Hurst index estimation and its application in Chinese stock market. *IEEE Access*, *9*, 93315-93330.
- Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Frontiers of Machine Learning and Finance. In M. F. Dixon, I. Halperin, & P. Bilokon (Éds.), *Machine Learning in Finance: From Theory to Practice* (p. 519-541). Springer International Publishing. https://doi.org/10.1007/978-3-030-41068-1_12
- Dufour, V., Pelé, M., Neumann, M., Thierry, B., & Call, J. (2009). Calculated reciprocity after all: Computation behind token transfers in orang-utans. *Biology Letters*, *5*(2), 172-175. <https://doi.org/10.1098/rsbl.2008.0644>
- Dunst, C., & Gorman, E. (2009). Development of Infant and Toddler Mark Making and Scribbling. *Centre for Early Learning Literacy Review*, *2*.
- Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all : Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184-194. <https://doi.org/10.1016/j.neuroimage.2016.10.001>
- Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological measurement*, *23*(1), R1.
- Elderfield, J. (1971). Drawing as suspended narrative. *Leonardo*, 13-22.
- Eom, C., Choi, S., Oh, G., & Jung, W.-S. (2008). Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets. *Physica A: statistical mechanics and its applications*, *387*(18), 4630-4636.
- Fernández, E., & Jelinek, H. F. (2001). Use of fractal theory in neuroscience : Methods, advantages, and potential problems. *Methods*, *24*(4), 309-321.
- Freeman, N. H. (1972). Process and product in children's drawing. *Perception*, *1*(2), 123-140.
- Freeman, N. H. (1993). *Drawing : Public instruments of representation*.
- Freeman, N. H., & Janikoun, R. (1972). Intellectual Realism in Children's Drawings of a Familiar Object with Distinctive Features. *Child Development*, *43*(3), 1116-1121. <https://doi.org/10.2307/1127668>
- Fukushima, K. (1980). Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193-202. <https://doi.org/10.1007/BF00344251>
- García-Diez, M., & Ochoa, B. (2020). Art origins: The emergence of graphic symbolism. *Encyclopedia of Global Archaeology*, 1-19.
- Gardner, H. (2017). Reflections on Artful Scribbles : The Significance of Children's Drawings. *Studies in Art Education*, *58*(2), 155-158. <https://doi.org/10.1080/00393541.2017.1292388>
- Gardner, H., & Wolf, D. (1987). The symbolic products of early childhood. In *Curiosity, imagination, and play : On the development of spontaneous cognitive motivational processes* (p. 305-325). Lawrence Erlbaum Associates, Inc.
- Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). *A Neural Algorithm of Artistic Style* (arXiv:1508.06576). arXiv. <http://arxiv.org/abs/1508.06576>

- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2022). *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness* (arXiv:1811.12231). arXiv. <https://doi.org/10.48550/arXiv.1811.12231>
- Glenny, R. W., Robertson, H. T., Yamashiro, S., & Bassingthwaite, J. B. (1991). Applications of fractal analysis to physiology. *Journal of Applied Physiology*, 70(6), 2351-2367.
- Gombrich, E. H., Hochberg, J., & Black, M. (1973). *Art, Perception, and Reality*. JHU Press.
- Goodenough, F. L. (1926). *Measurement of intelligence by drawings* (p. Pp. 177). World Book Co.
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples* (arXiv:1412.6572). arXiv. <https://doi.org/10.48550/arXiv.1412.6572>
- Grottanelli, V. L. (1961). Asonu Worship among the Nzema : A Study in Akan Art and Religion. *Africa*, 31(1), 46-60.
- Hagood, M. M. (2003). The Use of the Naglieri Draw-a-Person Test of Cognitive Development : A Study with Clinical and Research Implications for Art Therapists Working with Children. *Art Therapy*, 20(2), 67-76. <https://doi.org/10.1080/07421656.2003.10129396>
- Hanazuka, Y., Kurotori, H., Shimizu, M., & Midorikawa, A. (2019). The Effects of the Environment on the Drawings of an Extraordinarily Productive Orangutan (*Pongo pygmaeus*) Artist. *Frontiers in Psychology*, 10. <https://www.frontiersin.org/article/10.3389/fpsyg.2019.02050>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. 770-778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- Hearst, M. A., & Hirsh, H. (2000). AI's Greatest Trends and Controversies. *IEEE Intelligent Systems*.
- Henshilwood, C. S., d'Errico, F., Van Niekerk, K. L., Dayet, L., Queffelec, A., & Pollarolo, L. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature*, 562(7725), 115-118.
- Islam, M. A., Kowal, M., Esser, P., Jia, S., Ommer, B., Derpanis, K. G., & Bruce, N. (2021). *Shape or Texture : Understanding Discriminative Features in CNNs* (arXiv:2101.11604). arXiv. <https://doi.org/10.48550/arXiv.2101.11604>
- Jacob, G., Pramod, R. T., Katti, H., & Arun, S. P. (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-22078-3>
- Jolley, R. P. (2009). *Children and Pictures : Drawing and Understanding*. John Wiley & Sons.
- Joordens, J. C., d'Errico, F., Wesselingh, F. P., Munro, S., De Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J. R., & Kuiper, K. F. (2015). *Homo erectus at Trinil on Java used shells for tool production and engraving*. *Nature*, 518(7538), 228-231.
- Jumel, B. (2015). *Aide-mémoire-Dessin d'enfant-2e éd*. Dunod.
- Kawase, H., Okata, Y., & Ito, K. (2013). Role of huge geometric circular structures in the reproduction of a marine pufferfish. *Scientific reports*, 3(1), 1-5.

- Kim, B., Reif, E., Wattenberg, M., Bengio, S., & Mozer, M. C. (2021). Neural Networks Trained on Natural Scenes Exhibit Gestalt Closure. *Computational Brain & Behavior*, 4(3), 251-263. <https://doi.org/10.1007/s42113-021-00100-7>
- Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2018). Interpretability Beyond Feature Attribution : Quantitative Testing with Concept Activation Vectors (TCAV). *Proceedings of the 35th International Conference on Machine Learning*, 2668-2677. <https://proceedings.mlr.press/v80/kim18d.html>
- Kirsh, D. (1991). Foundations of AI : The big issues. *Artificial intelligence*, 47(1-3), 3-30.
- Kohts, N. (1935). *Infant ape and human child (instincts, emotions, play, habits)*.
- Krink, T., & Vollrath, F. (1997). Analysing spider web-building behaviour with rule-based simulations and genetic algorithms. *Journal of theoretical Biology*, 185(3), 321-331.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25. <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- Kubilius, J., Bracci, S., & Beeck, H. P. O. de. (2016). Deep Neural Networks as a Computational Model for Human Shape Sensitivity. *PLOS Computational Biology*, 12(4), e1004896. <https://doi.org/10.1371/journal.pcbi.1004896>
- Kubrick, S. (1968). *2001 : A space odyssey*.
- Kucak, D., Juricic, V., & Dambic, G. (2018). Machine Learning in Education—A Survey of Current Research Trends. In B. Katalinic (Éd.), *DAAAM Proceedings* (1^{re} éd., Vol. 1, p. 0406-0410). DAAAM International Vienna. <https://doi.org/10.2507/29th.daaam.proceedings.059>
- Ladygina-Kots, N. N. (2000). Infant Ape and Human Child. *Journal of Russian & East European Psychology*, 38(1), 5-78. <https://doi.org/10.2753/RPO1061-040538015>
- Lam, B. H. (2009). *The nature of drawing*.
- LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten Digit Recognition with a Back-Propagation Network. *Advances in Neural Information Processing Systems*, 2. https://proceedings.neurips.cc/paper_files/paper/1989/hash/53c3bce66e43be4f209556518c2fcb54-Abstract.html
- Lei, D., Chen, X., & Zhao, J. (2018). *Opening the black box of deep learning* (arXiv:1805.08355). arXiv. <https://doi.org/10.48550/arXiv.1805.08355>
- Lethmate, J. (1982). Tool-using skills of orang-utans. *Journal of Human Evolution*, 11(1), 49-64. [https://doi.org/10.1016/S0047-2484\(82\)80031-6](https://doi.org/10.1016/S0047-2484(82)80031-6)
- Lewis, O. J. (1969). The hominoid wrist joint. *American Journal of Physical Anthropology*, 30(2), 251-267. <https://doi.org/10.1002/ajpa.1330300211>
- Li, R., Lu, X., Li, S., Yang, H., Qiu, J., & Zhang, L. (2020). DLEP : A Deep Learning Model for Earthquake Prediction. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1-8. <https://doi.org/10.1109/IJCNN48605.2020.9207621>
- Liu, J., Li, Q., Yang, H., Han, Y., Jiang, S., & Chen, W. (2019). Sequence Fault Diagnosis for PEMFC Water Management Subsystem Using Deep Learning With t-SNE. *IEEE Access*, 7, 92009-92019. <https://doi.org/10.1109/ACCESS.2019.2927092>

- Longobardi, C. (2015). Reconsidering the scribbling stage of drawing: A new perspective on toddlers' representational processes. *Frontiers in Psychology*, 6, 9.
- Luquet, G.-H. (1913). Le problème des origines de l'art et l'art paléolithique. *Revue Philosophique de la France et de l'Étranger*, 75, 471-485.
- Luquet, G.-H. (1927). *Le dessin enfantin. (Bibliothèque de psychologie de l' enfant et de pédagogie.)*.
- Luquet, G.-H. (1920). *Les bonshommes têtards dans le dessin enfantin*.
- Lurçat, L. (1979). *L'Enfant et l'Espace : Le rôle du corps*. FeniXX.
- Ma, Y., Shi, W., Peng, C.-K., & Yang, A. C. (2018). Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep medicine reviews*, 37, 85-93.
- MacDonald, J. (2014). Alpha: The Figure in the Cage. *Relations. Beyond Anthropocentrism*, 2(2), 27-43.
- MacIntosh, A. J., Pelletier, L., Chiaradia, A., Kato, A., & Ropert-Coudert, Y. (2013). Temporal fractals in seabird foraging behaviour: Diving through the scales of time. *Scientific Reports*, 3(1), 1884.
- Mahendran, A., & Vedaldi, A. (2015). *Understanding Deep Image Representations by Inverting Them*. 5188-5196. https://openaccess.thecvf.com/content_cvpr_2015/html/Mahendran_Understanding_Deep_Image_2015_CVPR_paper.html
- Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *science*, 156(3775), 636-638.
- Martinet, L. (2018). *Dessine-moi un mouton: De nouvelles méthodes pour caractériser les dessins chez l'être humain (Homo sapiens) et le chimpanzé (pan troglodytes)* [These en préparation, Strasbourg]. <https://www.theses.fr/s269883>
- Martinet, L., & Pelé, M. (2021). Drawing in nonhuman primates: What we know and what remains to be investigated. *Journal of Comparative Psychology*, 135(2), 176-184. <https://doi.org/10.1037/com0000251>
- Martinet, L., Sueur, C., Hirata, S., Hosselet, J., Matsuzawa, T., & Pelé, M. (2021). New indices to characterize drawing behavior in humans (Homo sapiens) and chimpanzees (Pan troglodytes). *Scientific reports*, 11(1), Article 1. <https://doi.org/10.1038/s41598-021-83043-0>
- Martinet, L., Sueur, C., Matsuzawa, T., Hirata, S., Morimura, N., & Pelé, M. (2023). Tool assisted task on touchscreen: A case study on drawing behaviour in chimpanzees (Pan troglodytes). *Folia Primatologica*, 1(aop), 1-17. <https://doi.org/10.1163/14219980-bja10008>
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.
- McGillivray, A. (2022). On Childism. *Canadian Journal of Children's Rights / Revue Canadienne Des Droits Des Enfants*, 9(1), Article 1. <https://doi.org/10.22215/cjcr.v9i1.3942>
- Mendelowitz, D. M. (1967). *Drawing. (No Title)*.
- Mendes, N., Steinbeis, N., Bueno-Guerra, N., Call, J., & Singer, T. (2018). Preschool children and chimpanzees incur costs to watch punishment of antisocial others. *Nature Human Behaviour*, 2(1), Article 1. <https://doi.org/10.1038/s41562-017-0264-5>
- Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

- Monett, D., Lewis, C. W., & Thórisson, K. R. (2020). Introduction to the JAGI Special Issue 'On Defining Artificial Intelligence'-Commentaries and Author's Response. *Journal of Artificial General Intelligence*, 11(2), 1-100.
- Morris, D. (1962). The biology of art : A study of the picture-making behaviour of the great apes and its relationship to human art. (No Title). <https://cir.nii.ac.jp/crid/1130000795973714176>
- Pelé, M., Thomas, G., Liénard, A., Eguchi, N., Shimada, M., & Sueur, C. (2021). I wanna draw like you : Inter-and intra-individual differences in orang-utan drawings. *Animals*, 11(11), 3202. <https://doi.org/10.3390/ani11113202>
- Perez, H., & Tah, J. H. M. (2020). Improving the Accuracy of Convolutional Neural Networks by Identifying and Removing Outlier Images in Datasets Using t-SNE. *Mathematics*, 8(5), Article 5. <https://doi.org/10.3390/math8050662>
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*.
- Pinto, G., Accorti Gamannossi, B., & Ann Cameron, C. (2011). From scribbles to meanings : Social interaction in different cultures and the emergence of young children's early drawing. *Early Child Development and Care*, 181(4), 425-444. <https://doi.org/10.1080/03004430903442001>
- Pinto, W., & Smith, P. (1999). An artifact as history in art education. *Art Education*, 52(1), 19-24.
- Quaglia, R., & Saglione, G. (1976). *Il disegno infantile : Nuove linee interpretative*. <https://iris.unito.it/handle/2318/57636>
- Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks. *The Journal of Neuroscience*, 38(33), 7255-7269. <https://doi.org/10.1523/JNEUROSCI.0388-18.2018>
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347-1358. <https://doi.org/10.1056/NEJMra1814259>
- Ramm, A. (2005). What is drawing? Bringing the art into art therapy. *International Journal of Art Therapy*, 10(2), 63-77.
- Restoy, S., Martinet, L., Sueur, C., & Pelé, M. (2022). Draw yourself : How culture influences drawings by children between the ages of two and fifteen. *Frontiers in Psychology*, 13. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.940617>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). « Why Should I Trust You? » : Explaining the Predictions of Any Classifier (arXiv:1602.04938). arXiv. <http://arxiv.org/abs/1602.04938>
- Robert, A., & Roy, A. G. (1993). La modélisation fractale et la variabilité spatiale des phénomènes naturels. *Geographie physique et Quaternaire*, 47(1), 3-19.
- Rose, S. E., Jolley, R. P., & Burkitt, E. (2006). A review of children's, teachers' and parents' influences on children's drawing experience. *International Journal of Art & Design Education*, 25(3), 341-349.
- Rosenblatt, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/h0042519>
- Schiller, P. H. (1951). Figural preferences in the drawings of a chimpanzee. *Journal of Comparative and Physiological Psychology*, 44(2), 101-111. <https://doi.org/10.1037/h0053604>

- Seibert, D., Yamins, D., Ardila, D., Hong, H., DiCarlo, J. J., & Gardner, J. L. (2016). *A performance-optimized model of neural responses across the ventral visual stream* (p. 036475). bioRxiv. <https://doi.org/10.1101/036475>
- Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2017). *Grad-CAM: Why did you say that?* (arXiv:1611.07450). arXiv. <https://doi.org/10.48550/arXiv.1611.07450>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.
- Shelley, M. (1818). *Frankenstein, or the modern Prometheus*. London: Printed for Lackington, Hughes, Harding, Mayor & Jones.
- Shimada, M., & Lejys, V. (2021). Jeu, objets, outils et apprentissage chez les chimpanzés sauvages. *Techniques & Culture*, 76, 42-55.
- Simon, A. (2014). *La vérité sort-elle de la bouche des enfants? : Remarques épistémologiques sur les spécificités de l'enfance comme objet d'enquête en sciences sociales*.
- Smith, D. A. (1973). *Systematic study of chimpanzee drawing*. 406-414.
- Springenberg, J. T. (2016). *Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks* (arXiv:1511.06390). arXiv. <http://arxiv.org/abs/1511.06390>
- Sueur, C. (2011). A non-Lévy random walk in chacma baboons : What does it mean? *PloS one*, 6(1), e16131.
- Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology*, 10(3), 257-273. <https://doi.org/10.1007/s12194-017-0406-5>
- Svantesson, M., Olausson, H., Eklund, A., & Thordstein, M. (2023). Get a New Perspective on EEG : Convolutional Neural Network Encoders for Parametric t-SNE. *Brain Sciences*, 13(3), Article 3. <https://doi.org/10.3390/brainsci13030453>
- Theodorus, A., Nauta, M., & Seifert, C. (2020). Evaluating CNN interpretability on sketch classification. *Twelfth International Conference on Machine Vision (ICMV 2019)*, 11433, 475-482. <https://doi.org/10.1117/12.2559536>
- Turing, A. M. (2009). *Computing machinery and intelligence*. Springer.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 2579-2605.
- Vinter, A. (1999). How meaning modifies drawing behavior in children. *Child Development*, 70(1), 33-49.
- Viswanathan, G. M., Buldyrev, S. V., Havlin, S., Da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999). Optimizing the success of random searches. *nature*, 401(6756), 911-914.
- Wan, G. H. (2001). Changes in regional inequality in rural China : Decomposing the Gini index by income sources. *Australian Journal of Agricultural and Resource Economics*, 45(3), 361-381.
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1-37.
- Watanabe, S. (2012). Animal aesthetics from the perspective of comparative cognition. In *Emotions of animals and humans : Comparative perspectives* (p. 129-162). Springer.
- Waterson, R. H., Lander, E. S., Wilson, R. K., & The Chimpanzee Sequencing and Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), Article 7055. <https://doi.org/10.1038/nature04072>

- Weiss, B., Clemens, Z., Bódizs, R., Vágó, Z., & Halász, P. (2009). Spatio-temporal analysis of monofractal and multifractal properties of the human sleep EEG. *Journal of neuroscience methods*, 185(1), 116-124.
- Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-22244-7>
- Yamagata, K. (1997). Representational activity during mother-child interaction : The scribbling stage of drawing. *British Journal of Developmental Psychology*, 15(3), 355-366. <https://doi.org/10.1111/j.2044-835X.1997.tb00526.x>
- Yamagata, K., & Shimizu, M. (1997). *Development of constructive activity in early drawing*. 22-30.
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23), 8619-8624. <https://doi.org/10.1073/pnas.1403112111>
- Zeller, A. (2007). 'What's in a picture?' A comparison of drawings by apes and children. *2007(166)*, 181-214. <https://doi.org/10.1515/SEM.2007.056>

Annexes

I. Annexes de l'Article 2

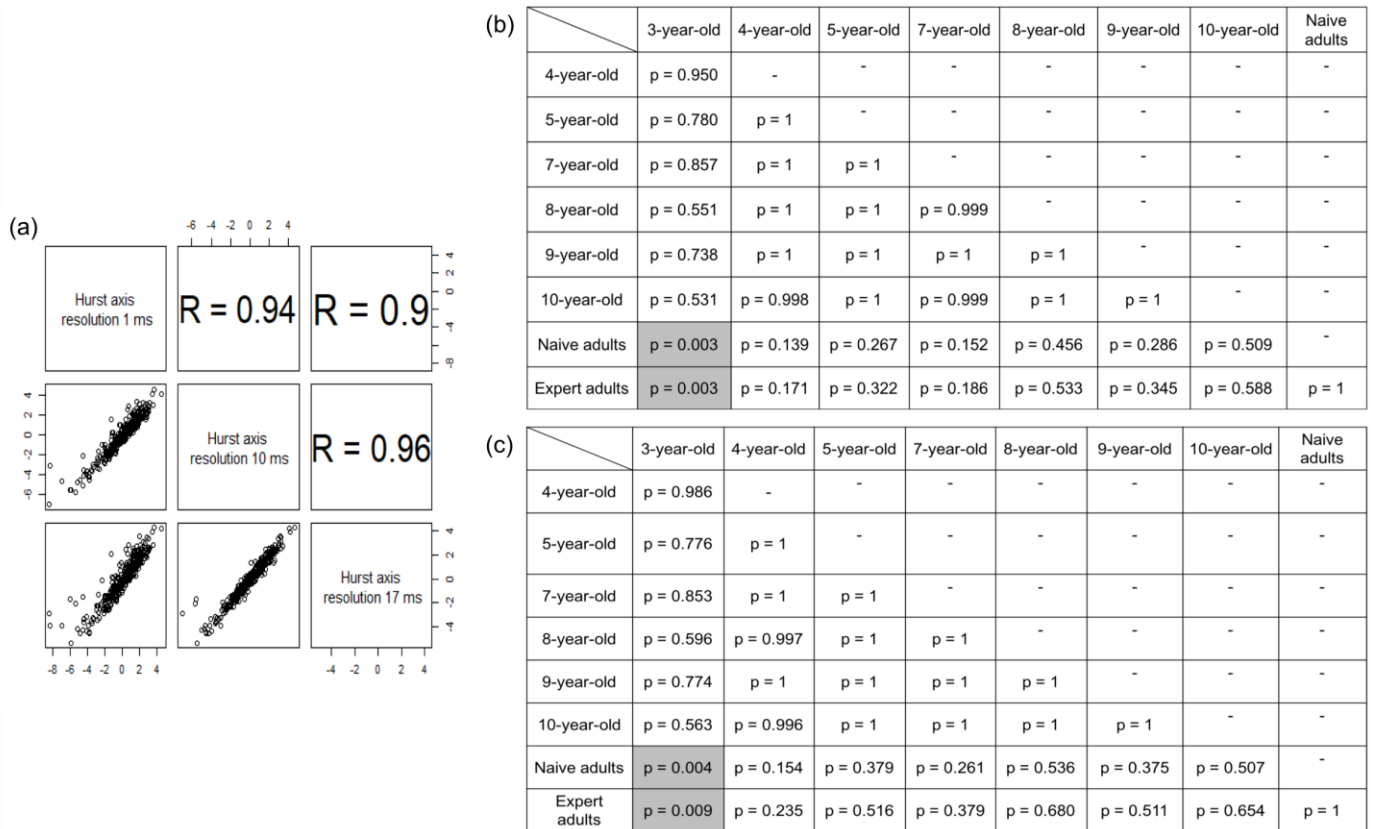


Figure 1. Analyses of different resolutions for the construction of the time series. (a) Pearson correlation matrix indicated the correlation between the Hurst axis values obtained with 3 different resolutions in the time series (1ms being the one used in the manuscript). (b) Pairwise comparisons of groups obtained from the selected model (GLMM, $df = 8$, $\chi^2 = 25.811$, $p = 0.001$) with a time resolution of 10 milliseconds. (c) Pairwise comparisons of groups obtained from the selected model (GLMM, $df = 8$, $\chi^2 = 23.23$, $p = 0.003$) with a time resolution of 17 milliseconds. The different resolutions give the same results with the 3-year-olds showing a higher value along the Hurst axis than adults, both, naïve and expert.

To understand the impact of removing a period of time in the sequences, and to ensure this will lead to consistent estimates, 50 drawings were randomly sampled. A chunk of 30 seconds was then randomly removed from the whole temporal sequence for each drawing, and the remaining parts were spliced. The 4 estimates were calculated on the first 50 seconds of the shortened sequences of drawings. The correlation of these estimates and the corresponding estimates on the first 50 seconds of the non-modified sequences were calculated. The correlations were strong (DFA: 97%, SWV: 96%, HAV: 92%, lowPSD: 94%), showing the consistency of this methodology and its robustness to missing segments of data.

Box 1. Explanation of simulations showing the robustness of fractal estimates to random removals of segments within the sequence.

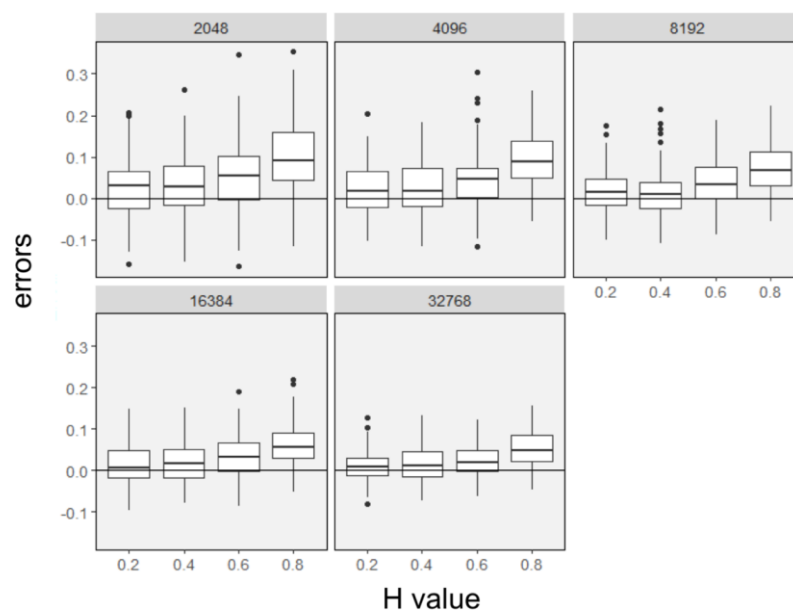


Figure 2. Estimation errors of the estimates of Monte Carlo simulations with the HAV method by varying H and the length of the time series.

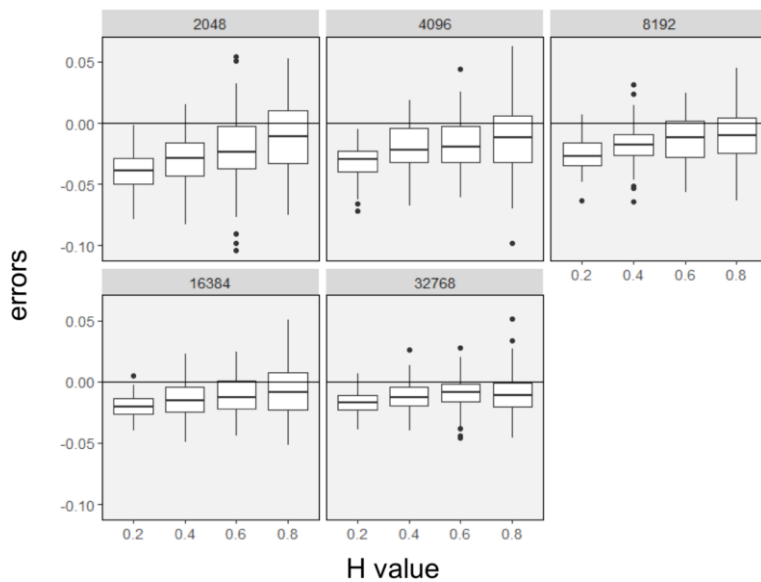


Figure 3. Estimation errors of Monte Carlo simulations with the SWV method by varying H and the length of the time series.

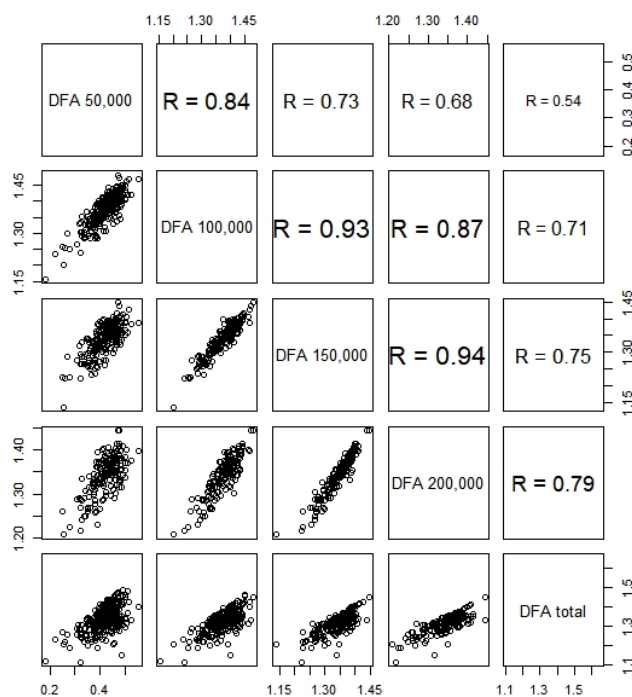


Figure 4. Correlations between DFA coefficients based on the first 50,000 points, 100,000 points, 150,000 points and 200,000 points and the total length of the times series.

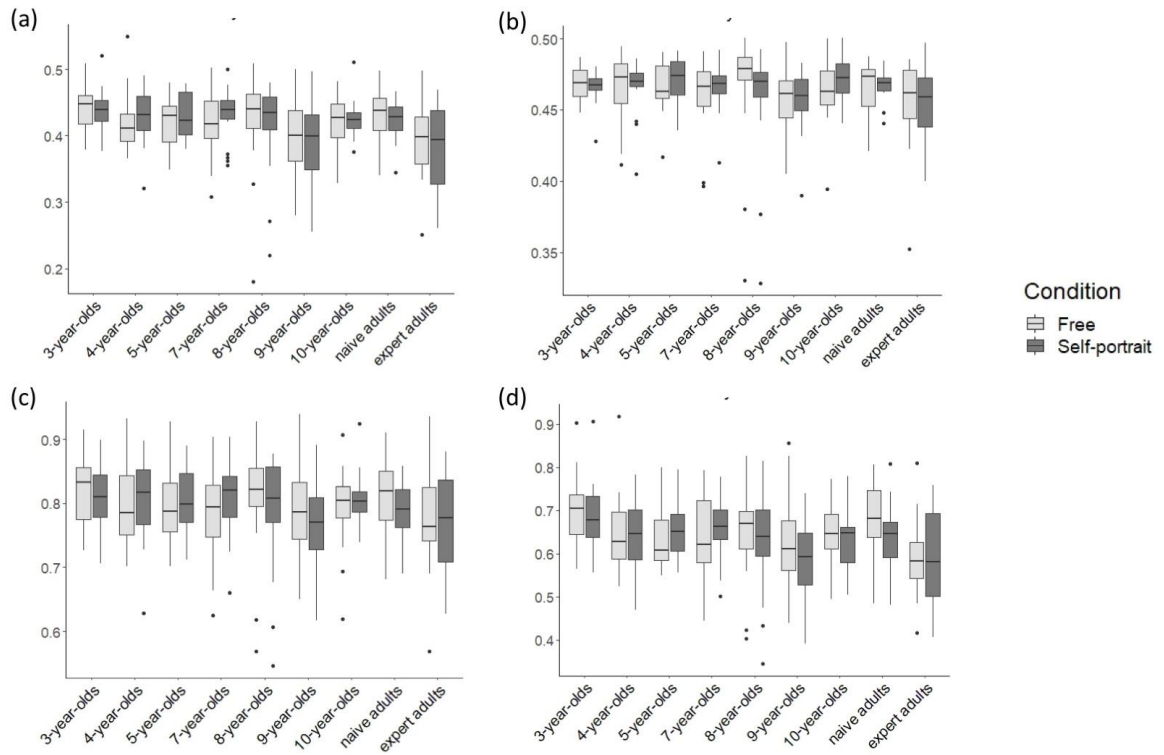


Figure 5. Boxplots of the Hurst estimates calculated with four different methods. (a) Hurst exponent estimated with (a) the Detrended Fluctuation Analysis (DFA), (b) the Power Spectral Density analysis (^{low}PSD), (c) the Scaled Windowed Variance (SWV) and (d) the Hurst Absolute Value method (HAV).

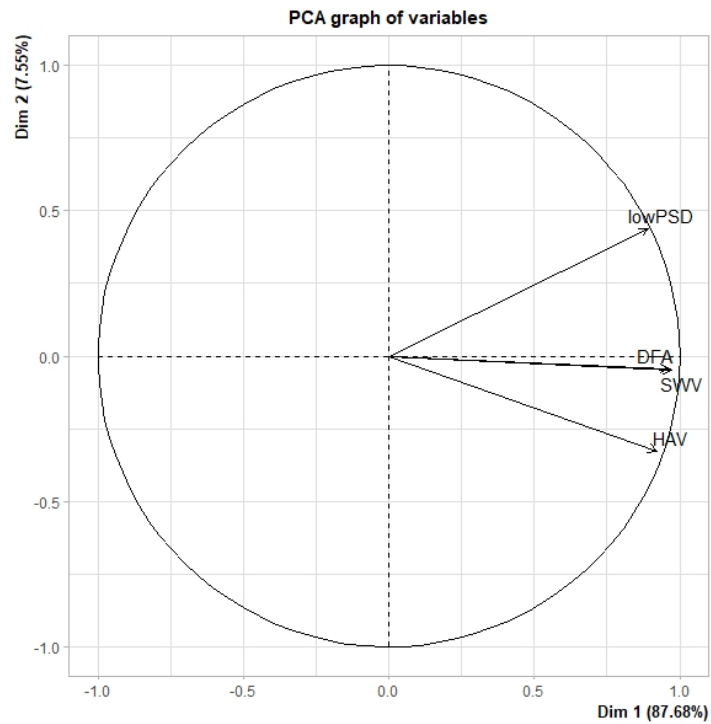


Figure 6. Correlation graph of the variables.

Table 1. Pairwise comparisons of groups obtained from the selected model (GLMM, $df = 8$, $\chi^2 = 22.842$, $p = 0.003$) made by averaging the estimates of H.

	3-year-old	4-year-old	5-year-old	7-year-old	8-year-old	9-year-old	10-year-old	Naive adults
4-year-old	$p = 0.950$	-	-	-	-	-	-	-
5-year-old	$p = 0.780$	$p = 1$	-	-	-	-	-	-
7-year-old	$p = 0.722$	$p = 1$	$p = 1$	-	-	-	-	-
8-year-old	$p = 0.832$	$p = 1$	$p = 1$	$p = 0.1$	-	-	-	-
9-year-old	$p = 0.701$	$p = 1$	$p = 1$	$p = 1$	$p = 1$	-	-	-
10-year-old	$p = 0.645$	$p = 0.999$	$p = 1$	$p = 0.1$	$p = 1$	$p = 1$	-	-
Naive adults	$p = 0.004$	$p = 0.139$	$p = 0.319$	$p = 0.273$	$p = 0.271$	$p = 0.332$	$p = 0.621$	-
Expert adults	$p = 0.010$	$p = 0.333$	$p = 0.05682$	$p = 0.186$	$p = 0.502$	$p = 0.587$	$p = 0.849$	$p = 1$

II. Annexes de l'Article 3

Supplementary material for Making drawings speak through mathematical metrics.

Cédric Sueur^{1,2}, Lison Martinet¹, Benjamin Beltzung¹, Marie Pelé³

1 : Université de Strasbourg, CNRS, IPHC UMR 7178,
Strasbourg, France 2 : Institut Universitaire de France,
Paris, France

3 : Anthro- Lab, ETHICS EA7446, Lille Catholic University, Lille, France

Corresponding author: Cédric Sueur, cedric.sueur@iphc.cnrs.fr ; IPHC
UMR 7178, 23 rue Becquerel 67087 Strasbourg, France

Details about metrics, their calculation and a value range for each instruction in
dataset 1

1) μ MLE

μ MLE is the maximum likelihood exponent that results from the analyses of the drawing-line length distribution. This analysis originates from the random walk or Lévy walk theory (Sueur, 2011; Sueur et al., 2011). The random walk analysis determines whether the distribution of drawing lines follows a power law or an exponential law. If the distribution follows an exponential law, we expect the drawing to be random, meaning that the individual who is drawing does not intend to represent any specific thing. In contrast, a power distribution should reflect a non-random and oriented behaviour, as found for the daily paths of animals in their natural environments (i.e. goal-oriented and efficient movements; Sueur, 2011 ; Sueur et al., 2011). Details about the use of this metric for drawing analysis are developed and detailed in (Martinet et al., 2021).

As the coordinate scoring of the drawing was continuous (one point per frame), we focused on active changes (Byrne et al., 2009; Noser and Byrne, 2014): a selection of points was carried out for each drawing via a change-point test under R software (version 1.1.383; CPT, script available in Noser and Byrne, 2014). Two consecutive points (i and j) in the drawing determined a step or a vector of a length $L(i, j)$. We then calculated the step lengths S on Excel with latitude x and longitude y (in pixels). Step lengths between 0 and 10 pixels were removed since they often corresponded to very short, inactive movements such as imprecise lines or finger sideslips, and caused inaccuracies. We then determined whether the step length frequency distribution of a drawing followed a power law ($y = a \cdot x^\mu$) or an exponential law ($y = a \cdot e^{-x/\lambda}$) using the Maximum Likelihood Method (Edwards et al., 2007; Martinet et al., 2021). Log-likelihood of the exponential and power distributions for each drawing could then be compared using the Akaike Information Criterion (AIC). We retained the model retained (power or exponential) with the lowest AIC, with a minimum difference of 2 between the two AICs (Burnham and Anderson, 2004). All the drawings produced followed a power law. The Maximum Likelihood Estimate of the power law exponent μ_{MLE} was then used to draw conclusions on the efficiency of the representation for each drawing. This index is comprised of values between 1 and 3. The higher the index, the more the line was considered to be directed, well planned and efficient (Bartumeus et al., 2005; Viswanathan et al., 1996).

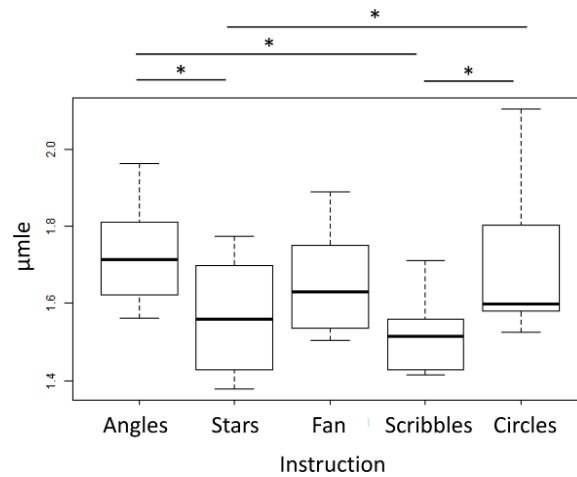


Figure S1. Boxplot of the μMLE according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test gives a $p=0.0003$.

2) Drawing distance

Drawing distance is the total distance of drawing in pixels, from the first point to the last. Two consecutive points (i and j) in the drawing determined a step or a vector of a length L (i, j). We then calculated the step lengths S on Excel with latitude x and longitude y (in pixels). We calculate the sum of all lengths S as the drawing distance.

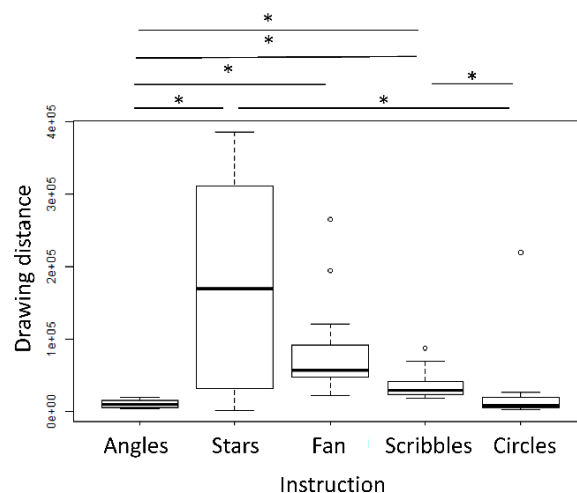


Figure S2. Boxplot of the drawing distance according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. Kruskal-Wallis test gives a p -value < 0.0001 .

3) Angle distribution metric

Goal directedness in animals is assessed by individuals moving in straight lines with limited tortuosity, except when they arrive at a food resource site and start to forage. Drawing intentionality can be considered similar to animal food research efficiency (Martinet et al., 2021; Martinet and Pelé, 2020). Analyses of angle distributions (i.e. turning angles) between two trajectories or lengths can be considered as a reliable way to measure goal directedness and tortuosity (Bartumeus et al., 2008; Hurford, 2009; Potts et al., 2018). A turning angle is the difference in direction for two successive vectors or steps. We followed the same methodology as described in 1. μ MLE. As coordinate scoring of the drawing was continuous (one point per frame), we focused on active changes (Byrne et al., 2009; Noser and Byrne, 2014): a selection of points was carried out for each drawing via a change-point test under R software (version 1.1.383; CPT). Step lengths between 0 and 10 pixels were removed as they often corresponded to very short, inactive movements such as imprecise lines or finger sideslips, and caused inaccuracies. Function ATAN2 in Excel was then used to calculate the angle (in radians) between two consecutive points and convert radians into degrees. We applied corrections to only retain angles between 0° and 180° , as movements are oriented. We then calculated the survival distribution of angles (i.e. going from 1 or 100% of points to 0) for each drawing. A cubic function was fitted to the distribution. Examples of drawings and the respective distributions are given in Figure S3.

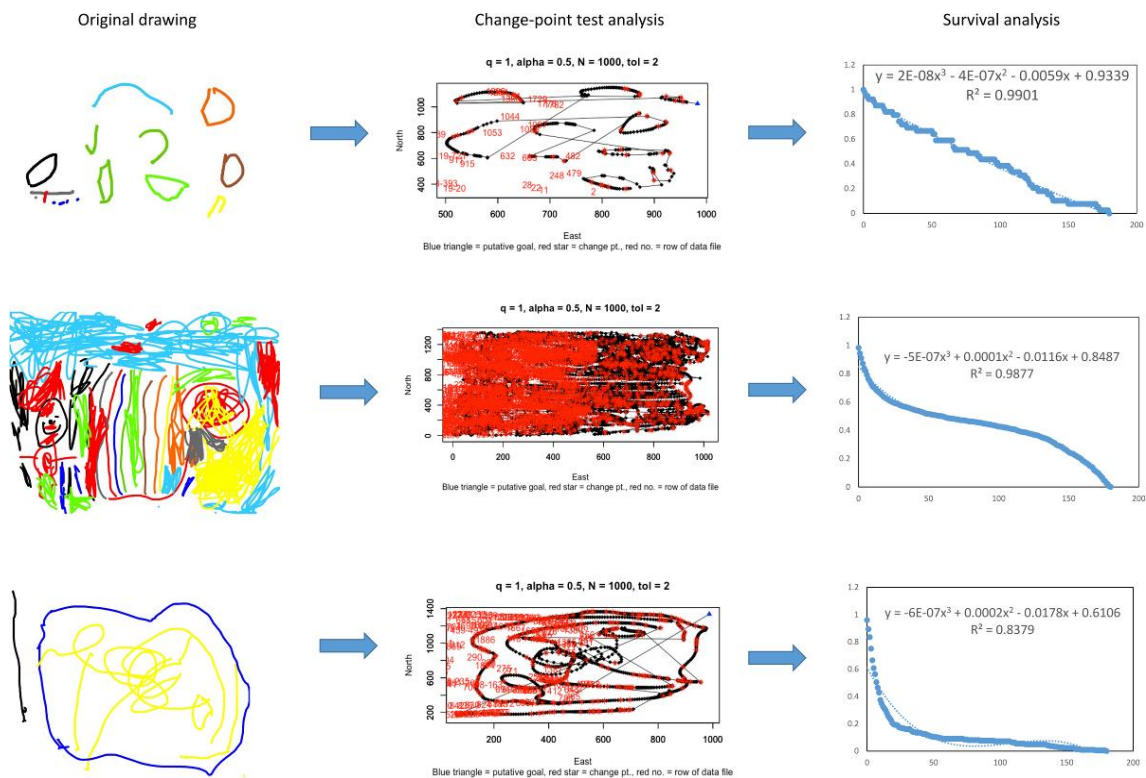


Figure S3. Analysis of angles distribution. The original drawing is transformed into vectors (i.e. trajectories) following the change-point test, and angles are calculated. The angle survival distribution is then fitted with a cubic function.

The cubic function is $y = -ax^3 + bx^2 - cx + d$. The more the distribution looks like a straight line, the lower the values of a , b and c will be. The more the curve looks like a sigmoid or inversed power function, the higher the values of a , b and c will be. The three constants a , b and c are highly correlated and can be combined using a principal component analysis, where one dimension explained 82% of variance for dataset 1 (dimension 2 = 15%) and 90% of variance for dataset 2 (dimension 2 = 5%). We used the values of dimension 1 as values for the angle distribution metric for each drawing.

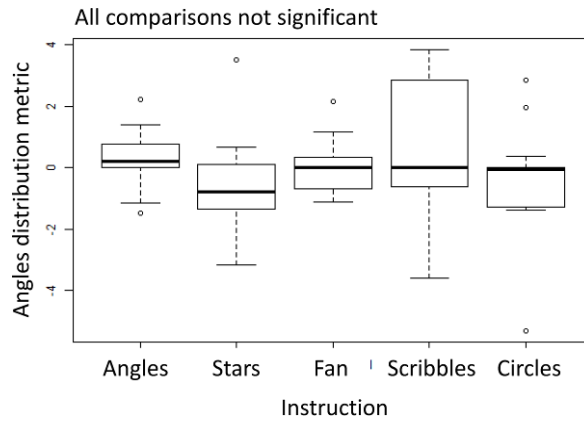


Figure S4. Boxplot of the angle distribution metric according to each instruction. ANOVA testing provides a P value of 0.329

4) Minimum convex polygon

The minimum convex polygon draws the smallest polygon around points with all interior angles measuring less than 180 degrees. Minimum convex polygons are common estimators of home range (Nilsen et al., 2008) but represent the cover of drawing on the screen, between 0 (no drawing at all) to 100% (the drawing covered the entire screen). We used the scissors select tool in GIMP 2.10.22 (Lecarme and Delvare, 2013; Peck, 2006) to select the minimum convex polygon of drawings.

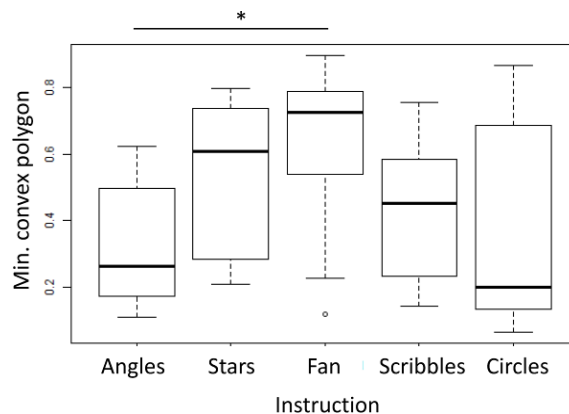


Figure S5. Boxplot of the minimum convex polygon according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA testing provides a P value of 0.016.

5) Hurst index

Details about the calculation of this metric can be found in Beltzung et al. (2021). Fractality or long-term processes can be measured by different methodological approaches, each of which has its own fractal statistical parameter. Here, the difficulty lies in the fact that numerous estimators have been defined for each parameter, yet the effectiveness of these very estimators is still debated in the literature (Stadnitski, 2012; Stadnytska et al., 2010). Studies often focus on one or a small number of estimators without a rigorous reason (such as comparing them). As a consequence, there is no simple and systematic way to estimate the long-memory process, which often results in errors or misleading conclusions in studies (Karagiannis et al., 2006). The most widespread way to assess and quantify long-memory processes in temporal sequences is the estimation of the Hurst exponent H . Indeed, a behavioural state is influenced by previous states following two different scenarios: *persistence* ($H > 0.5$) when a positive correlation occurs, meaning that a long sequence is likely to be followed by a long sequence in the future, and *anti-persistence* ($H < 0.5$) when a long sequence is likely to be followed by a short sequence, i.e. a negative correlation (Delignières et al., 2005). Here, we combined different methods using a PCA as explained in Beltzung et al. (2021).

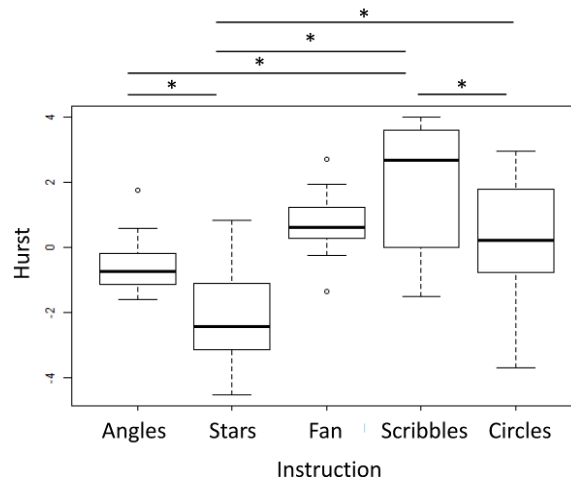


Figure S6. Boxplot of the Hurst index according to each instruction. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test gives a $p < 0.0001$.

6) Temporal Gini index

The Gini index was calculated on the binary sequences (drawing / non-drawing) and describes the uniformity of the distribution. This index is a real number between 0 and 1, where a value of 0 indicates perfect equality, and a value of 1 indicates maximal inequality. The interpretation of this index on behavioural sequences can be questioned, as well as the temporal sequence to consider. The Gini index was calculated by using the R package “DescTools” (Signorell et al., 2016).

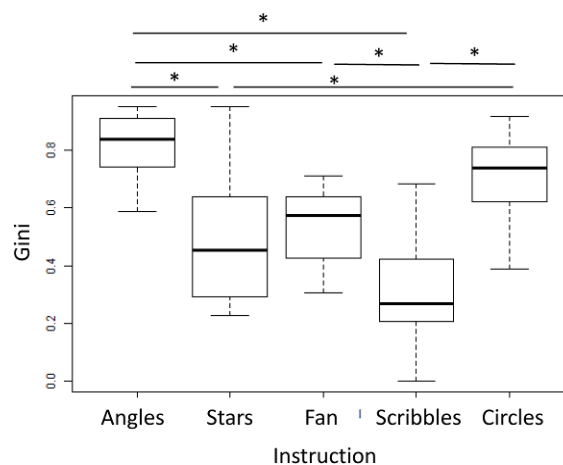


Figure S7. Boxplot of the Gini metric according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p < 0.0001$.

7) Entropy index

The Shannon entropy index (Gray, 2011; Leff, 2007) has not been calculated for temporal binary sequences, but for the cumulative sum of drawing time at each second. This index characterises the quantity of information contained in a variable. The higher the quantity of information, the higher the uncertainty (the entropy). For a string of sequences with n distinct sequences, each sequence has a frequency of p . The entropy index was calculated by using the R package “DescTools” (Signorell et al., 2016). The entropy of Shannon H is calculated according to the formula:

$$H = - \sum_{i=1}^n p_i \log_2 p_i$$

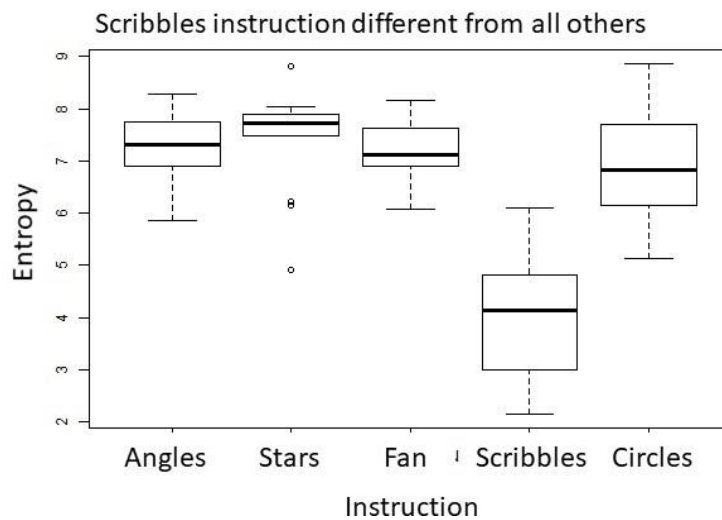


Figure S8. Boxplot of the entropy metric according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p < 0.0001$.

8) Drawing duration

The duration of the drawing session is defined as the time from the first point on the touchscreen to the last, including the time spent drawing and time without drawing between these two points.

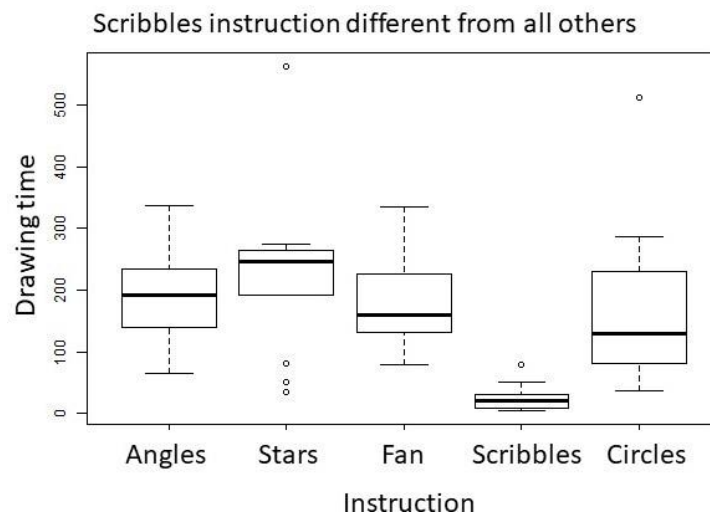


Figure S9. Boxplot of the drawing duration according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p < 0.0001$.

9) Number of sequences

Number of drawing and non-drawing sequences during the test.

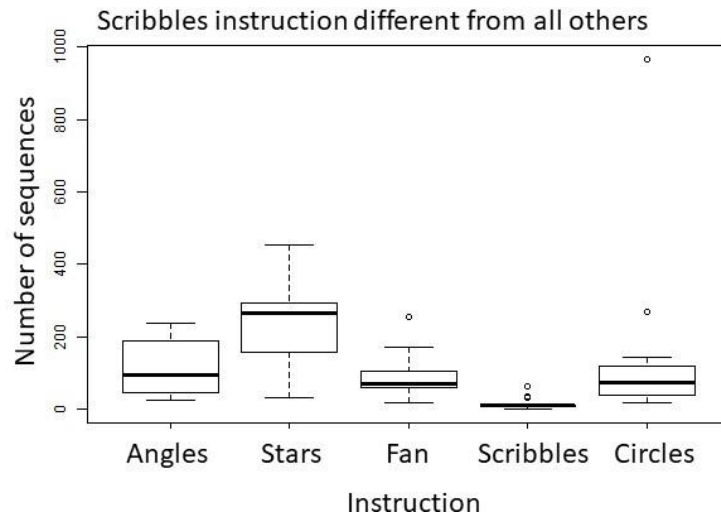


Figure S10. Boxplot of the number of sequences according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. Kruskal-Wallis test: $p < 0.0001$.

10) Drawing speed

Drawing distance divided by the drawing duration (i.e. duration of drawing only, not including duration of non-drawing during the test).

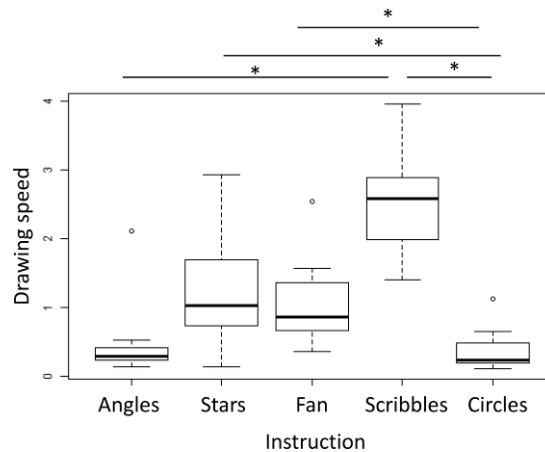


Figure S11. Boxplot of the drawing speed according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p < 0.0001$.

11) Drawing time proportion

Duration of drawing divided by the duration of drawing session.

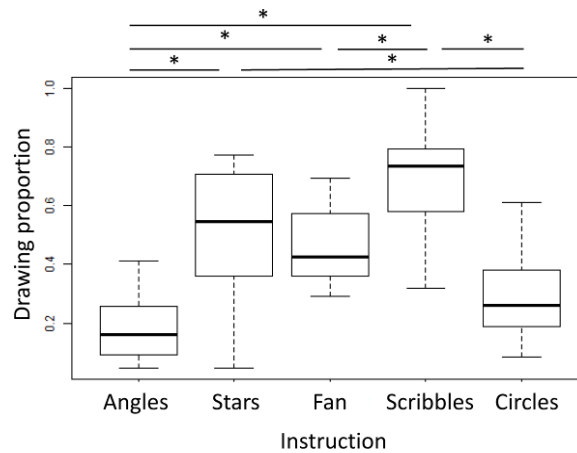


Figure S12. Boxplot of the drawing proportion according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p < 0.0001$.

12) Mean colorimetric profile

Mean distribution of intensity levels for the Red, Green, or Blue colours respectively and after removal of the white (screen) colour on the parts covered by drawing (after selection with the minimum convex polygon).

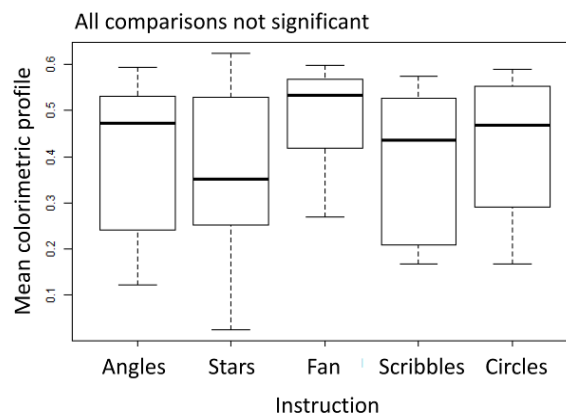


Figure S13. Boxplot of the mean colorimetric profile according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p = 0.319$.

13) Standard deviation of the colorimetric profile

Standard deviation of the distribution of intensity levels for the Red, Green, or Blue colours respectively, on the parts covered by drawing (after selection with the minimum convex polygon).

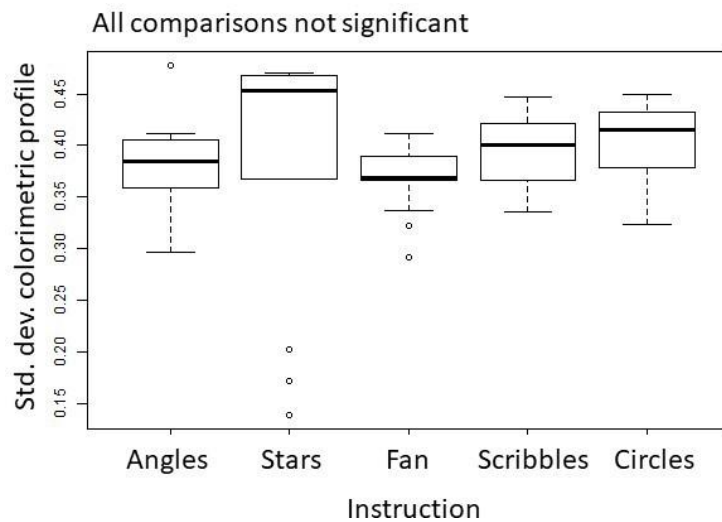


Figure S14. Boxplot of the standard deviation of the colorimetric profile according to instructions. Stars indicate significant difference between two instructions with $p < 0.05$. ANOVA test: $p = 0.675$

14) Number of colours

Number of colours used from the ten proposed colours. The ten proposed colours are black, grey, red, blue, dark green, light green, sky blue, brown, orange and yellow.

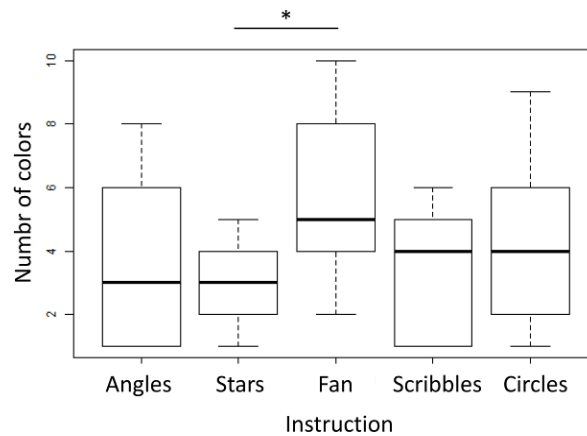


Figure S14. Boxplot of the number of colours according to instructions. Stars indicates significant difference between two instructions with $p < 0.05$. Kruskal-Wallis test: $p=0.04$.

Supplementary figures and tables referenced in the main text

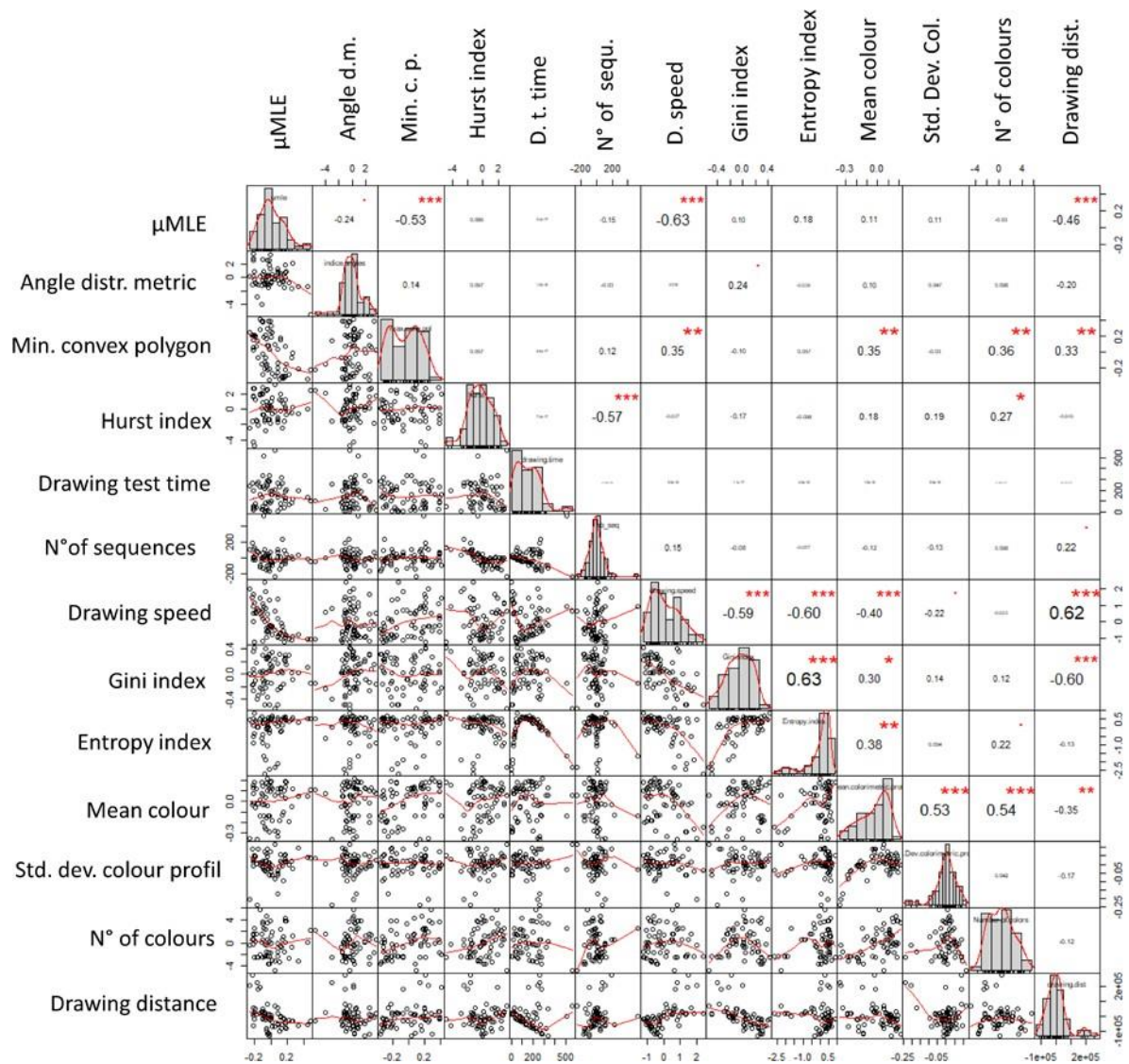


Figure S15. correlation chart of the 13 metrics for dataset 1 (first step, after removal of the drawing-time proportion and correction according to the drawing test time). The diagonal of the graph provides the distribution of each metric, whilst the bottom left and the top-right provide the correlation figure and the correlation coefficient between two metrics, respectively. Statistical value is given with the correlation coefficient: * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

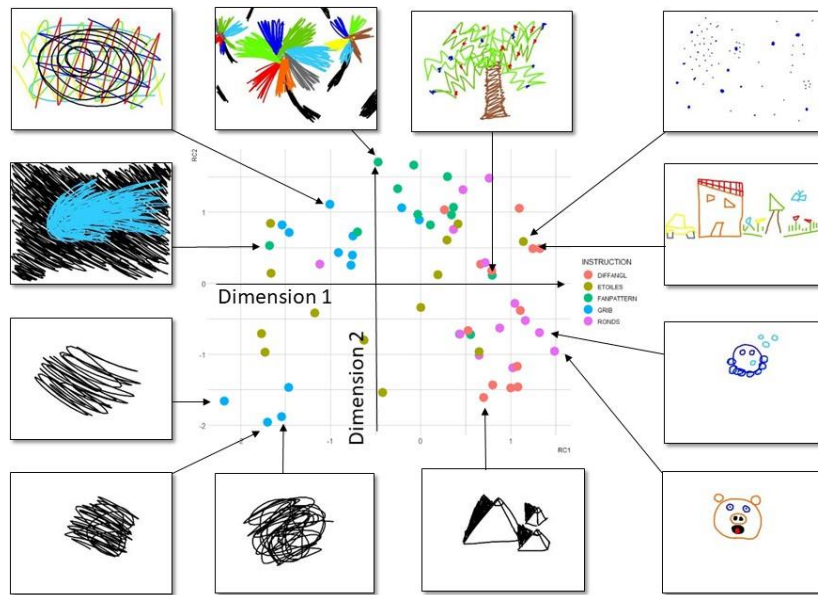


Figure S16a. Examples of Dataset 1 (first step) drawings according to Dimension 1 and Dimension 2, provided by the PCA. Dimension 1 may represent representativeness in drawing whilst Dimension 2 may represent diversity in drawings.

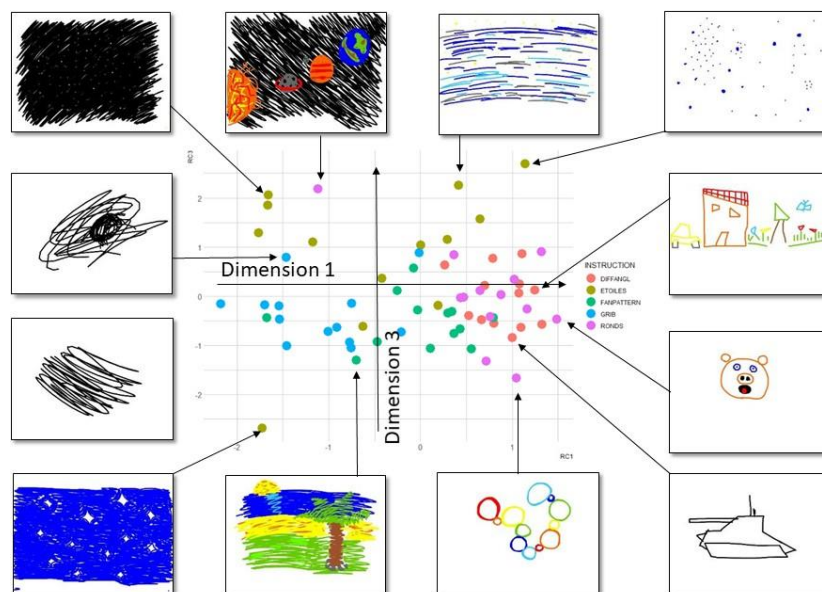


Figure S16b. Examples of Dataset 1 (first step) drawings according to Dimension 1 and Dimension 3, provided by the PCA. Dimension 1 may represent representativeness in drawing whilst Dimension 3 may represent periodicity in drawings.

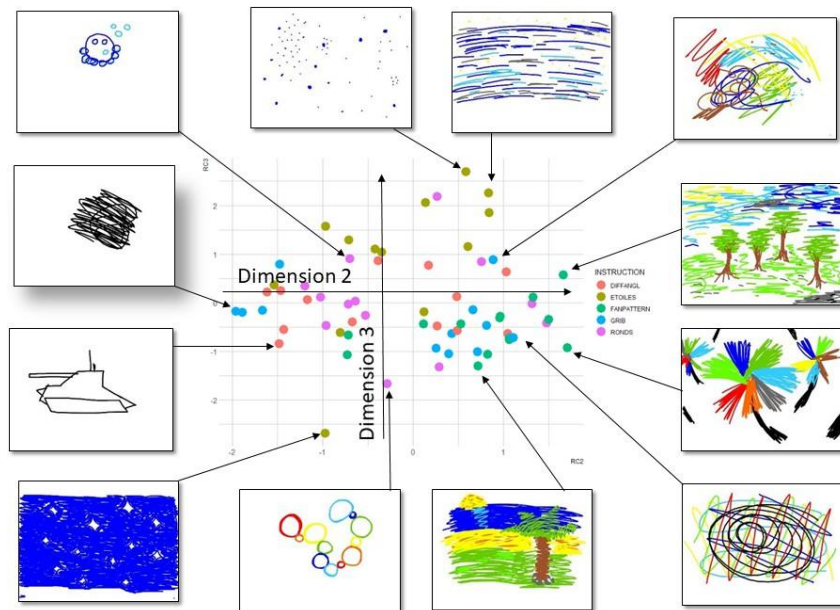


Figure S16c. Examples of Dataset 1 (first step) drawings according to Dimension 2 Dimension 3, provided by the PCA. Dimension 2 may represent diversity in drawing whilst Dimension 3 may represent periodicity in drawings.

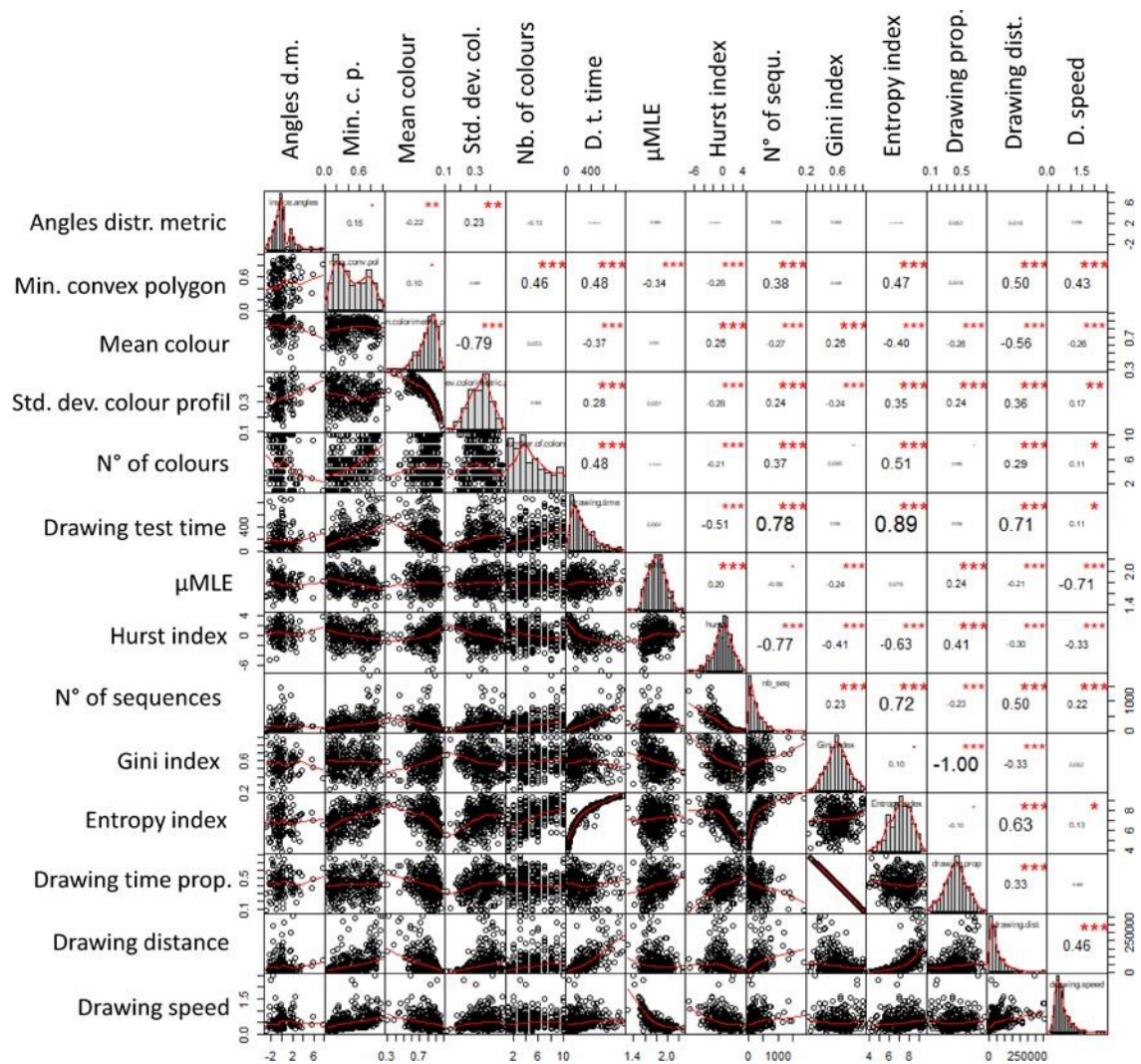


Figure S17. Correlation chart of the 14 metrics for dataset 2 (second step). The diagonal of the graph provides the distribution of each metric, whilst the bottom left and the top-right provide the correlation figure and the correlation coefficient between two metrics, respectively. Statistical value is given with the correlation coefficient: * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

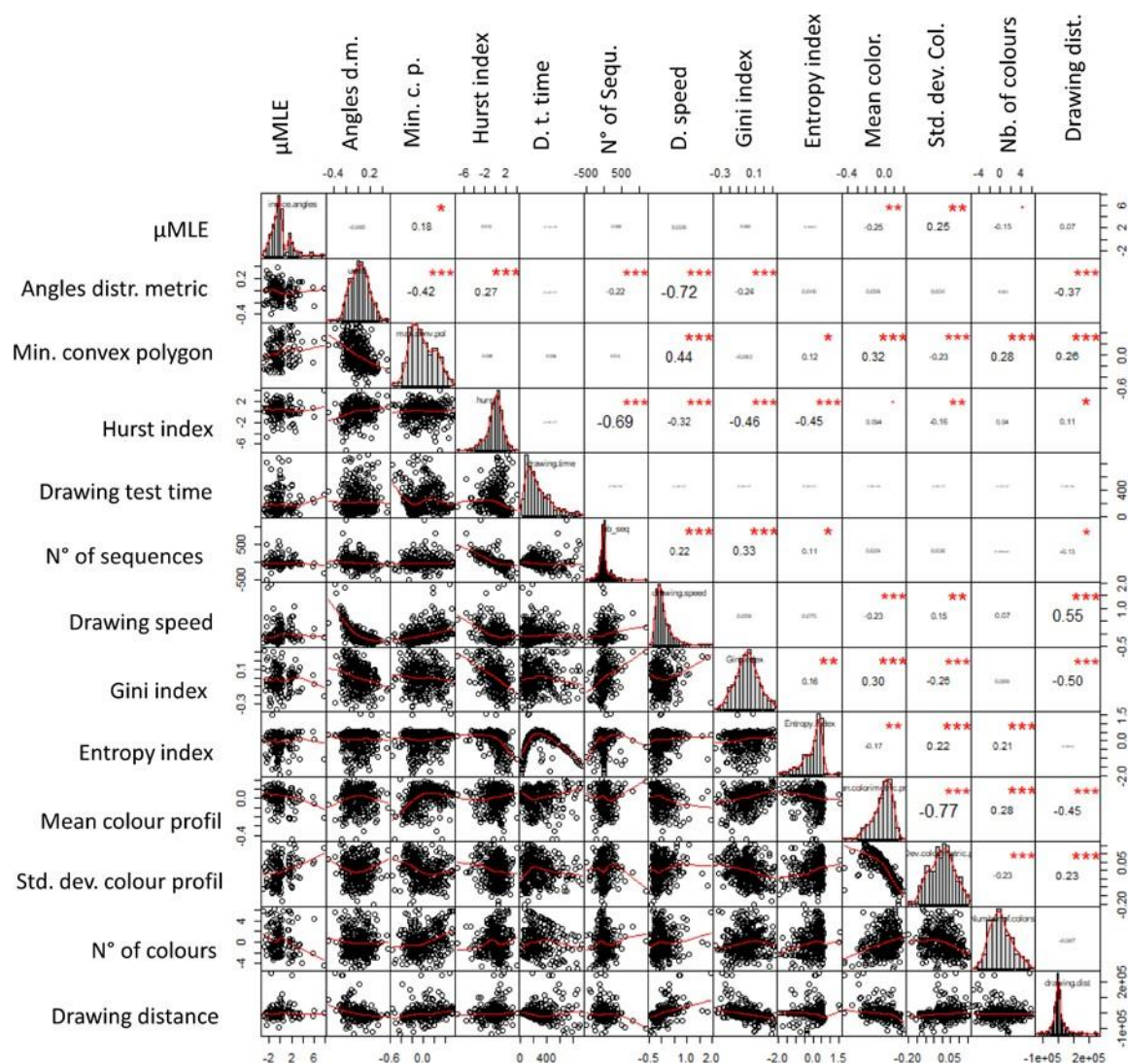


Figure S18. Correlation chart of the 13 metrics for dataset 2 (second step, after removal of the drawing-time proportion and correction according to the drawing test time). The diagonal of the graph provides the distribution of each metric, whilst the bottom left and the top-right provide the correlation figure and the correlation coefficient between two metrics, respectively. Statistical value is given with the correlation coefficient: * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

Table S1. Loading of the 13 metrics on the three Varimax rotation PCA dimensions (before selection of variables with loading superior to 0.4) of Dataset 1

	Dim1	Dim2	Dim3
μ MLE	0.639	-0.373	0.314
Angle distribution metric		0.347	
Min. conv. pol.	-0.451	0.725	-0.159
Hurst index	-0.14	0.195	0.849
Drawing test time			
Number of sequences	-0.115		-0.755
Drawing speed	-0.911		-0.12
Gini index	0.762	0.301	-0.213
Entropy index	0.624	0.375	-0.235
Mean colour profile	0.365	0.758	0.254
Std. dev. colour profile	0.221	0.296	0.398
Number of colours		0.716	
Drawing distance	-0.746		-0.228

Table S2. Mann-Whitney test for each metric of each dimension between the scribbles of toddlers and those drawn by adults. *p*-values < 0.05 are in bold print.

Dimension	Metrics	W	p-value	Mean for toddlers	Mean for adults
1	μ MLE	101	0.225	-0.07±0.14	-0.12±0.09
	drawing distance	47	0.097	30239±72983	26505±14882
	drawing speed	29	0.007	0.31±0.73	1.04±0.68
2	Mean colour Profil	93	0.473	0.01±0.09	-0.05±0.16
	Number of colours	118	0.03	2.79±2.68	0.09±1.
3	Hurst	84	0.769	1.29±1.03	0.71±2.0
	Number of sequences	11	<0.0001	-52.12±33.0	5.39±18.3

Utilisation de réseaux de neurones convolutifs pour mieux comprendre l'évolution et le développement du comportement de dessin chez les Hominidés

Résumé

L'étude du comportement de dessin peut s'avérer très informative cognitivement et psychologiquement, tant chez les humains que chez les autres primates. Cependant, cette richesse d'information peut également être un frein à son analyse et à son interprétation, en particulier en l'absence d'explication ou verbalisation de son auteur. En effet, il est possible que l'interprétation donnée par un adulte d'un dessin ne soit pas en accord avec l'intention première du dessinateur. Durant ma thèse, j'ai montré que, bien que généralement considérés comme des boîtes noires, les réseaux de neurones convolutifs (CNNs) peuvent permettre de mieux comprendre le comportement de dessin. Dans un premier lieu, l'utilisation d'un CNN a permis de classer des dessins d'une femelle orang-outan selon leur saison de production ainsi que de mettre en avant une variation de style et de contenu. De plus, une approche ontogénique a permis de quantifier la similarité entre des productions de différents groupes d'âges. Par la suite, des modèles plus interprétables ainsi que l'application de nouvelles méthodes d'interprétabilité pourraient permettre de mieux déchiffrer le comportement de dessin.

Mots-clés : Intelligence artificielle, apprentissage profond, interprétabilité, comportement de dessin, primates

Summary

The study of drawing behavior can be highly informative, both cognitively and psychologically, in humans and other primates. However, this wealth of information can also be a challenge to analysis and interpretation, particularly in the absence of explanation or verbalization by the author of the drawing. Indeed, an adult's interpretation of a drawing may not be in line with the artist's original intention. During my thesis, I showed that, although generally regarded as black boxes, convolutional neural networks (CNNs) can provide a better understanding of the drawing behavior. Firstly, by using a CNN to classify drawings of a female orangutan according to their season of production, and highlighting variation in style and content. In addition, an ontogenetic approach was considered to quantify the similarity between productions from different age groups. In the future, more interpretable models and the application of new interpretability methods could be applied to better decipher drawing behavior.

Key-words : Artificial intelligence, deep learning, interpretability, drawing behavior, primates